Implicitly Coded Knowledge:
Content-Based Representations of Image Sequences

by

Chee Kong Mok

S.B. Physics
Massachusetts Institute of Technology
Cambridge, Massachusetts
1985

SUBMITTED TO THE MEDIA ARTS AND SCIENCES SECTION, SCHOOL OF ARCHITECTURE AND PLANNING, IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF

MASTER OF SCIENCE IN VISUAL STUDIES

AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1991

Signature of the Author
....
Chee Kong Mok
Media Arts and Sciences Section
January 18, 1991

Certified by
....
Walter Bender
Principal Scientist, Media Laboratory
Thesis Advisor

Accepted by
....
Stephen A. Benton
Chairman
Departmental Committee on Graduate Students

Implicitly Coded Knowledge:
Content-Based Representations of Image Sequences

by

Chee Kong Mok

Submitted to the Media Arts and Sciences Section, School of Architecture and Planning, on January 18, 1991 in partial fulfillment of the requirements of the degree of Master of Science in Visual Studies at the Massachusetts Institute of Technology

Abstract

In this implementation, content is encoded in an image sequence by implicitly coding segmentation information in the rearranging of the look-up-table. Context is encoded by links between sequences. The idea of content-based representations is defended by claiming that content is already inherently interwoven into representations. A discussion is made also about the reading of the image in the framework of an interactive movie.

Thesis Advisor: Walter Bender
Title: Principal Scientist, Media Laboratory

This work was supported in part by International Business Machines Corporation.

Contents

Preface

The danger of subjective description is that it can color the content of the described, leading to bias in the manner a subsequent reader will be informed. There is thus a desire to achieve as much an objective description as possible, to allow the reader enough latitude to interpret on his own. This has been a silent idealism fueling the technology of imagery.

However, it is a strange paradox that as technology presents us with ever increasing power for the manufacture and circulation of imagery, through photographic and now digital images, these formats of imagery themselves tend increasingly to hold less semantic information. Attention to content has been sacrificed in the search for objectivity. This is an undesirable stance. The first chapter, **Digital Messages**, argues for the futility of the search for objective descriptions of images. The describer must accept the subjectivity inherent in the task of providing a description of an image; he has thus henceforth dispensed with any allusions of definitiveness or universality. He is finally allowed to provide what appears to him the most appropriate description from his vantage point. In other words, we have no choice but to entrust him with deciding the contents. This then leads to the importance of developing content-based representations.

Before we begin with the task of creating an example of content-based representations, we should first look for the history of such representations, not confining our domain to the particular field of digital imagery. We must examine the paradigms upon which these representations arose; we must examine the very field of *representation*, its implications and any introspective reflections. This will provide insight into problems we are likely to face, and be forewarned of the implications and limitations of our task. This discussion is covered in the second chapter, **Image Representations**, of this paper.

After these two chapters, we are ready to discuss the specific representation of digital image sequences that has been implemented in concert with this paper. The method chosen is one whose *content* includes knowledge of cognitive objects in the image. The method is a simple one, a slight variation on the traditional representation of digital images (that of a rectangular image intensity array), with the addition of encoding content implicitly within the look-up-table by rearranging entries. All this is outlined in the third chapter on **Methodology**.

As will be demonstrated, an indirect but desired consequence of this encoding is the forming of predicates, breaking the tight bonds of the pixel and frame as fundamental units. This is a turn away from the mounting division of imagery through pixelization in recent history beginning with Mondriaan. This division and search for a fundamental unit is a result of the desire to define images precisely to enable exact communication. Predicating (grouping of pixels or yet other predicates) forms higher order organization of images and maintains and even magnifies the manipulability of digital images.

The first three chapters speak of writing images, and thereafter we proceed with the issue of its subsequent reading. How is the information that has been encoded accessed? Representing imagery with higher orders of manipulability and organization presents new possibilities for displaying such images. The ease in decoding and access of the encoded content offered by the method implemented suggests an interactive format for reading. The fourth chapter **The Interactive Movie**, speaks about interacting with the *moving image*[1].

[1] *moving image* and *image sequence* will be used interchangebly.

Interactive systems transform the nature of reading and writing itself. The impact, meaning and consequences of interactivity are examined. The many pittfalls and problems, both computational and sociological, are also examined to understand why they arise. Possible solutions are suggested.

We end with **Currents** which will recapitulate certain main points, and reevaluate what has been achieved, reexamine what the original goals and intentions were. The implementation and its implications are defended, and its shortcomings discussed and resolved.

To summarize, the purpose of the following discussions and inquiries is not so much to lead to an ideal representation of moving imagery, but to touch on some issues that might encourage us to further examine the problems at hand. By partially removing ourselves from direct coding, moving toward some more general questions about representations, we begin to see parallels in related fields that may yield additional insights. We desire to discover patterns of study, and based on these results, place image coding in the context of ordering of information for the purpose of deciphering knowledge.

Chapter 1

Digital Messages

Digital Images

A digital image, in its conventional format (and we speak now not of synthesized images, but of *registrations*[1] of the visible world), like a photograph, is a thing: it is an object. But through this object, the digital image conveys meanings: it emits messages. The digital image is coded with knowledge, explicitly through culture.

We shall attempt to extract the object from the nest of messages that accompanies it and, by studying this object, examine any remnant murmurs of messages that derive purely from the very format of representation. We are looking for *digital messages*, knowledge that is implicitly coded in digital images.

[1] *Registration* as a precise type of *recording*.

1.1 Cultural Messages

A digital image has messages. Culture plays a role in producing these messages.

Barthes[Barthes 77] extensively examines and speaks at length about messages and meanings in images. Though his direct focus is photographic images, his ideas can be read more generally. We will take the liberty to tailor and extend these ideas with the goal of informing ourselves of the nature of digital imagery.

What are these *cultural* messages?

Documentation

A digital image is a document. A digital image, like any *registered* image, or a record of any form, functions often as a document, existing for the purpose of reading and re-reading, for the (re)examination and study of the registered. In documenting, messages are formed.

In the creation of this image as document, there is, beside the purely mechanical process of registration, the processes of emission and reception[Barthes 85]. Through the very manner of the processes of writing (emission) and reading (reception), messages are constructed. The first of thiese messages is the literal message: it is the pure identification of the represented scene.

Reception

Outside of this first message (the literal message), a second message emerges from the impact of the image. The second message emerges as a function of reading, and readings differ widely according to the context of an image's exhibition and according to its readers. The second message is transformed by the adjacency of other materials, by the timing of its showcase, and by the beliefs and previous exposures of its audience.

The message that is imparted depends on its reception.

Emission

This second message that is imparted also depends on its emission.

Its "maker", in the process of framing, in his choice of subject and his treatment (lighting, photogeny etc.), and in his use of objects (props, minor or major), forms a specific *recording* of the scene. The image is, through these processes, *coded* with signs, and these significations, producing connotations of the image, draw upon conventions of the particular culture. The success and effectiveness of this second message require that its readers maintain access to this codebook of knowledge, this collection of conventions. In a particular culture, a bookcase may signify intellectuallism, for example, or a recently extinguished candle may signify the ephemeral, or lighting from below may signify unknown horrors. The codebook is known.

Culture

We see then that the reception and emission of an image is a function of sociology. This observation, in turn, forces the study of the structure of these messages—the processes of their manufacture and consumption—to be necessarily entwined with a study of social structure.

The study of images is a study of culture.

The study is an historical one as the message rests upon a set of culturally *accumulated* signifiers; its signs and expressions are pregnant with particular associations by virtue of the practices in a given society.

The study of the image at this level is a study of these conventions themselves, their origins, their impact and their nature.

1.2 Digital Image as Analog to Depicted

A digital image is also an object.

We have spoken thus far about connoted messages deriving from culture. In this section, we wish to examine the *digital image as an object*, outside of the image's emission and reception, to study its autonomous structure, purged of sociology and its complexities, to be able to examine the formal skeleton on which the object rests. We wish to see how the foundations of the object form, guide, and limit the object's secondary and tertiary meanings. What is left of the message if we must not consider its reading within the cultural context?

1.2.1 Analog as Object

What is the image as an object?

A *digital image as an object*, in its conventional format, like a photograph, speaks of visual reality[2], echoing it as a perfect "copy". It is a map to a territory, the scene that is *registered.* And even if this map, this image, is not visual reality itself, it is at least its perfect analogue. Barthes[Barthes 85] calls it an *analogon.* It is the pure and simple denotation of reality.

It is curious that the etymology of the word *image*, which links it to the Latin *imitari*, from which *imitation* also derives. The Webster dictionary defines image as

[2]Visual reality as distinct from dimensional reality.

image, *n.* Latin *imitari* to imitate. 1: a reproduction or imitation of the form of a person or thing.

The *analogon* is an object. The *analogon* is an autonomous part of the image. As an object, it is free and independent—it lives of itself. It is indescribable, for, as like any description, it cannot be redescribed, for the act or redescription transforms the object. The image as an object is resistant to interpretation, resistant to interrogation, lest interrogation disintegrates it, as a mummy decays when "interrogated."

1.2.2 Object in Construction

How is this *analogon* constructed?

The mechanical process of *registration*, by which a scene is recorded, is a transfer, not a transformation; it cannot be called a transformation in that no codes are produced to guide its subsequent reading. The process can, however, be termed a transfer, in that the semblance is transferred from the depicted scene to its depiction. In this transfer, there is no need for codes. The digital image, in its mechanical form, that is, the *analogon*, is a pure correspondence to the reality it records. It is true that, as in any imaging system, artifacts are undoubtedly introduced—spatial and intensity resolution, for example. But, aside from this, within these limitations, the *analogon* is a direct mapping.

The physical process of registration in and of itself imparts no codes to the registered image.

The *analogon* is international; it is non-denominational.

1.3 Significations of the Digital Image

1.3.1 Mythical Non-signification of the Object

We have separated the *analogon* of the digital image from significations deriving from culture. Is this object then objective, clear of any significations? It is, at first glance, seemingly so. Noting that all second messages are cultural, and that such messages could not exist without a book of codes, then it follows that since the *analogon* has no book, it is then pure and void of references to codes: the *analogon* remains, as it was when constructed, a pure denotation. This section aims to inspect this situation.

These questions are new. They are brought about by technology which has presented the very possibility of the *analogon.* To begin this inspection, we start with the study of the production of images before and after the advent of technology that has brought about the *analogon.*

1.3.2 Signifying before Technology

In the reproductive disciplines before photography, before the rapid advent of technology, in the mimetic arts such as drawing or painting, not every visual detail was "copied", and often very little. Thus these arts were, compared with photography, much less denotative of the depicted scene. Nevertheless they did not cease to impart strong messages.

Imperfections offer Second Messages

These representations created semblances of the *analogon.* The *image as object*, the *analogon*-semblance contained "mistakes", visual inaccuracies. Because of these *mistakes*, the image as object, in the function of denoting the depicted scene, cannot perform without distraction.

These *mistakes* located the difference between the *seen* and the *said*, and these "locations" are exactly the "space" for connotations, for significations, for the coding of *second* messages. The lack of exactness of these representations resulted in the corruption of denotation: connotations and denotations were intertwined.

Messages were formed through the very *style* of reproduction. Representation, the image as object, incorporated connotations. The image as object was cultural.

In the recording of the depicted scene onto the object as representation, regulated transformations (perspective for example, or expressionist strokes), produced *connoted* messages, using references that are common in the community and generally known by the intended readers. The image is read intact. Outside of this community, however, where the appropriate book of codes is absent, the connoted messages are lost.

Thus, the physical processes before the advent of technology inherently produced codes. These codes mapped the image to the represented, and bridged the reader with the image.

Imperfections Insist on Second Messages

The act of rendering reflected the general disposition of the mind. A reproduction was thus more than a pure denotation of the depicted scene itself; it included what the scene appeared to be to the depictor's senses. Aristotle notes that the renderer often imitated things not as they are, but as *they ought to be.* These arts are then less representation of outer reality than they are of subjective expressions.

Even in the event of a meticulous rendering, imitating the scene in an "objective" fashion with the conscious goal of objectivity, however successful it is, the production of messages persists.

There was a *signature* of objectivity. There is *connoted* objectivity.

Before technology, all representations of images were attached to signifiers. There are signifiers outside of those produced by the processes of reception and emission. The representation itself (the image as object), signifies.

Before technology, representations (images as objects) signified.

1.3.3 Signifying after Technology

Then came technology. Photography brought with it the era of *production*: copies of descriptions are easily manufactured. This is followed by digital images, with the era of *process*: copies are not only produced but the copies were no longer distinct. All copies are identical. There is no longer an original.

Perfection is Cultural

Technology offers us a new order of mimetic representations. Photography, followed by digital imagery, presents us with so perfect an analog of the depicted, the *analogon*, that it requires no code. The only "knowledge" that is needed to read such images is the perceptual. Apparently there are no cultural messages, no second messages.

We take a bolder step presently in pursuing a further analysis of the the *analogon*. It is claimed here that the apparent is misleading and that, outside of the aforesaid cultural messages deriving from emission and reception, there are indeed second messages deriving from the very manner of construction of the representation, messages which are therefore embedded into the *analogon* itself.

In the next section, we will excavate these connoted messages from the *analogon*, unearthing the connotations of technology.

1.4 Connotations of Technology

1.4.1 The Impossibility of the Non-Connotative

Why do we believe that the *analogon* is purely denotative, that it is an *objective* representation of reality?

The human "maker" of a digital image has no intention of making the image not objective. On the contrary, it is often one of his conscious goals to make it objective. The maker detaches himself from the process of making, creating the machinery that takes over the function of image creation. This registering machinery does not "think", does not know the codes of culture. It is incapable of connoting and is thus, supposedly within the reach of objectivity.

But this "objectivity" risks being mythical for, although the "maker" is removed from the process of connoting, the technology that has brought it about is *cultural.* The human, or the society that he is part of, is ultimately the maker of the image. Hidden within the process of registering—this process of transcribing reality—are cultural messages. It is interesting to examine what these *unintended* cultural messages are, for in in doing so, we may perhaps might bring to the surface hints of our belief system.

Objectivity cannot be achieved. Pure denotation is impossible.

1.4.2 Phobia of the Incomplete, Scorn for the Inexact

Why do we want to achieve objectivity? Why do we believe in, and have a desire for, *objective* descriptions? Forming some understanding of these questions might help us discover these claimed "connotations of technology" and to put them into perspective.

It is sanctified that every individual reads for himself. This mythical demand is held supreme. To grant the possibility of this, the writing must be exact, it must be complete, free of connotations, for any connotation leads the reader—we must not read for others, lest we mislead them. Hence grows the suspicion and phobia for subjective descriptions.

But, given that we have the limitation of being finite with our descriptions, it is not possible to have a complete description for any general image, only if it was because that we do not know to what granularity to choose to describe, and the innumerable types of information the reader might request. A complete and exact description is impossible.

Despite obvious problems, the search for an objective description continued within its known limitations. What was sought after was a description that was precise enough in itself, so that a re-description would result in an exact copy of the first description. The image would then transcend its manifestation as an object. The image returns to become a concept. Piet Mondriaan [Mondriaan 20] was amongst the first to attempt to formulate a formal system, beginning to create exact definitions for spatial and color units—a pixelization of images. Digital images of today achieve many of these desired properties. Exact copies can be made because the description of the "original" is *completely* precise, enabling complete transmission of an image across a communication channel, on the other end of which an identical image can be reconstructed. This surreal achievement is applauded.

Pixelization of images has its achievements, but it has not achieved objectivity.

1.4.3 Intensity is the Message, Adjacency is the Order

Finally, if digital images are subjective, what then *are* some examples of their embedded connotations of technology?

There is always an order that we impose on our interaction with the natural world so as to enable us to have a reference with which to make possible controlled perception and enable us to accumulate knowledge. This is the *episteme* of a culture: a criteria for ordering.

In the digital image, as well as in the photograph, there is a direct correspondence in the spatial dimension, and a direct proportionality in intensity, between the image and the represented scene. In this instance then, the criteria, which we have no reason to believe are not subjective, upon which the description is sown, number two—these being, firstly, locality and adjacency and, secondly, light intensity.

The former, that of locality and adjacency, might be precariously defended as following natural laws and relationships which exist free from distortion by personal affect. Criteria that use these properties, that seemingly exist independently of mind, can perhaps be justifibly proclaimed as objective.[3] After all, do not natural interactions between matter, interactions such as gravity, seem localized and to be a function of the *euclidean*[4], or the *adjacent* in any geometric space?

Likewise, the latter criteria, that of light intensity, may too be argued and propped up. But these arguments are not strong enough to make us not conclude that embedded within the *analogon* are precisely these two cultural epistemes, these criteria for ordering, for knowing. The acceptance of the *image plane* as a good representation of the visual element is undoubtedly, as will be further demonstrated in the next chapter, cultural.

[3]This topic will be dealt with in detail in a later chapter.

[4]The distance between points in Euclidean geometry.

The digital image is not hermetic to the epistemes of society. The *analogon* reflects these epistemes. The *analogon* contains connotations of the beliefs of society, specifically of society's technology.

1.5 Technology of Images

1.5.1 Limitations of Technology

The *analogon* is not purely denotative; the *image as object* implicitly reflects cultural epistemes.

We have observed the futility of objective descriptions of digital imagery: all representations contain some decided information. Despite all attempts at objectivity, some criteria has been chosen, consciously or otherwise, with which information is described. Certain types of information are included in the description.

In realizing this futility of the search for objectivity, we salvage what we can, for not all is in despair; whether the ideal of objectivity was actually so desirable had not even been discussed. At this point, in accepting the inherent subjectivity in the task of describing images, it is important to decide what the most guides and motivations are. We must set ourselves to define new goals.

One major focus should be on the consideration of exactly what types of information we wish to encode in our "subjective" descriptions. Not believing in the sanctity of intensity-adjacency descriptions, frees us to rethink other appropriate representations of images. We must entrust the describer with deciding the *contents* of the representation. The idea of representations that are content-based becomes primary. And these *contents* is what the describer says the image *contains.*

In the run for objectivity, technology of image representations had progressed to bring various techniques of desirable and undesirable qualities and capabilities. We can now re-examine these qualities and take advantage of the strengths of technology for the fulfillment of new goals. What is the range and type of information that is available, obtainable, codable, and what are the techniques and implications of such developments?

Within the limitations of technology, what can it offer us?

1.5.2 Offerings of Technology

It is a strange paradox, that while technology has enabled the increase the circulation of information (it is easier to make copies of the photograph *analogon* than it is of paintings), the content of the photograph *analogon* compared to, say, painting, is correspondingly diminished. Attention to content has been sacrificed in the search for objectivity. This is an undesirable situation and it need not be for there is no inherent reasons why technology would bring methods that void content.

On the contrary, there is a gamut of properties of digital imagery that can be harnessed to good effect. In the tradition of Cubism and Perspective[Panofsky], we can consider the *analogon* of the digital image to be an array of symbols upon which we can perform transformations. The *analogon* is ultimately manipulable. Any particular section of the *analogon*, a unit of data (and there are a large range of definitions of *unit*), can be addressed. The digital image has a virtual presence: it is manifested in the visual but it exists really in the space of memory banks and mathematical matrices. This space lends itself operational to the realms of combinatory algebra, and transformable to all systems of equivalence. Thus the digital image, aside from its manifested visual, is a meta-image in the sense that it can be trivially manifested in many forms. Its manipulability, and its consequent ductility, allows much of any transformations to be implementable. Arbitrary transformations on the image can be performed, that set the image into formats that bring out information otherwise obscured. Within these innumerable formats, *content* can either be hidden and incorporated, or better yet, formats can be defined with *content* as its base.

All digital data (image, sound, etc.) have a common fundamental unit of *bit*, and thus begin to blur the distinctions between different media, making multi-media not a fancy but almost the default. Conversion from one medium to another, *transcoding*, from text to sound for example, becomes unnecessary. Text and sound become derivable from the same data. Text and sound are but different *readings* of the same data. "Content" remains constant.

Epilogue, and Preamble to next chapter

Epilogue

In reading, different readers deduce different knowledge and have divergent intentions. In writing, as in *re*presenting, the author intends that the material is read more than once—either by himself or by other parties to whom he imparts a message. That a writing cannot avoid being but one specific interpretation, imply that any representation cannot meet the desires and demands of all its readers. Accepting that there is no hermeneutical neutrality in any representation, suggests that the format of representations should not linger on attempts to be universal, but that it should convey its own particular message clearly. The only way the reader is capable of achieving his own intentions is that he subscribes to many multiple particular descriptions.

Preamble to next chapter

To focus on content-based representations of imagery, we must begin by putting the inquiry into the context of representations in general. We can then find precedents of representations of this type, not having to confine ourselves to examples within the field of imagery. While examining these precedents, we must closely consider the very field of *representation*, its reflections and implications. With this behind us, we can perhaps have better insights into the problems we are likely to face, and be forewarned of the limitations and harvests of our tasks. The next chapter begins with this discussion of image representations.

Chapter 2

Image Representations

In the last chapter, **Digital Messages**, we discussed the fall of a cultural paradigm, that of objectivity. The quest for objectivity is no longer pursued in any field, be it quantum theory or post-modernist architecture. This quest is replaced by a new percept—that of *subjective knowledge* with a *fabric of interconnecting relationships.* We are interested, here, in the sorts of representations that arise from this percept. What sorts of representations can express and promote the interests of this new manner?

It is the intentions of this paper to examine a representation that might incorporate and serve this new thought. It is in the next chapter, **Methodology**, that we demonstrate the implementation of a *content-based representation* where we formulate content-units which are linked by context-units.

So as to enable the next chapter to be read in the proper framework, our purpose here, in this chapter, is two-pronged. In the first section, we will examine how representation is inextricably linked with knowledge; how the types of representation reflect cultural beliefs, desires and understanding. The second section will take us through a focussed survey of digital image representations that are of relevance to our case, both in the sense of historical and conceptual developments. In this survey, we will observe that the form of digital image representations has headed toward a structuralist mode (that of structures and establishment of relationships between different hierarchical units of data), and onto a post-structuralist stance (that of multiplicity and inter-(cross)relationships).

2.1 Knowledge and representation

Nietzsche: "There are no facts, only interpretations."

Knowledge and representation are interwoven. The way we understand things determines how we describe them. The way we describe things determines how we will further understand them.

2.1.1 Representing

Representation is the technique by which notes are made and from which ideas are formed. In divergent fields of inquiry, issues of representation are of supreme importance, for they often determine the very success or failure of a quest. We note that the term *representation* has seemingly specific meanings in different fields. These meanings, at first glance, seem unrelated.

As Blatt [Blatt 84] notes about representing:

> "in painting it refers to technical procedures for recreating a segment of nature on a 2-D surface, in psychological theory the term refers to a construction of a sense, image, or conception of a segment of nature in one's mind, and in science the term refers to a model of a physical structure".

Without ignoring conventions peculiar to specific fields, *representation* refers to a process of symbol construction[Blatt 84], and in this sense the term is equivalent across fields, despite practical differences in the media in which symbol construction occurs.

Representation is a manifestation of the paradigms of inquiry, which in effect determines the very kinds of possible knowledge. It is not suprising that representations across different fields are structurally similar in a particular culture, as they express, are limited by and in turn limit, the culture's general conceptual schemes for the interpretation of the universe. Foucault[Foucault 74] calls these schemes *cultural epistemes.*

2.1.2 Classifications

Classifications reflect how we understand and order things.

Representation begins with classifying. A primary goal of representation, and of human knowledge in general, is to establish some ordering of information, from which we can "spin a web of understanding". This science of ordering is seen as a matter of sorting, dividing and enveloping—giving birth to schemes of <u>classification</u>. In this act of classifying, we delimit the *Same* to distinguish it from the *Other*, and thus begin to "tame the wild profusion of existing things"[Foucault 74], managing (and henceforth manipulating) the otherwise incomprehensible splattering of things and events.

We cannot represent without knowlege, and there is no knowledge without classification.

Borges[Borges 57] notes that a certain Chinese encyclopedia reports that animals are divided into

1. belonging to Emperor

2. embalmed
3. tame
4. sucking pigs
5. sirens
6. fabulous
7. stray dogs
8. included in the present classification
9. frenzied
10. innumerable
11. drawn with a very fine camel-hair brush
12. et cetera
13. having just broken the water pitcher
14. that from a long way off look like flies

Foucault says of this:

> In the wonderment of this taxonomy, the thing we apprehend in one great leap, ... is the limitation of our own [system of thought], the stark impossibility of thinking *that*. ... Each of these strange categories can be assigned a precise meaning and a demonstrable content; some of them do certainly involve fantastic entities—fabulous animals and sirens—but, precisely because it puts them into categories of their own, the Chinese encyclopaedia localizes their power of contagion; it distinguishes carefully between the very real animals (those that are frenzied or have just broken the water pitcher) and those that reside solely in the realm of imagination.

Classifications echo the criteria with which ordering is performed. Classifications yield *signifiers* with the smallest units of classes as the signified. With the assembly of these signifiers, higher classes are generated. To illustrate, we see that a *dog* is an instance of a (*furry*)-(*animal*).

Representation begins with classifying; it begins with the construction of symbols. *Representation reflects knowledge.*

2.1.3 Cultural Codes

The choices of representation reflect the desires and beliefs of culture.

It has been argued that classifications and representations reflect the *Weltanschauung* of their period or culture. Panofsky [Panofsky] does this in his discussion of the representation of space in art as a symbolic form. He shows that a culture's conception of nature is expressed in its rendition of objects and space. Italian Renaissance's linear perspective, he demonstrates, echoes their view of space as infinite, homogenous and isotropic.

The forms of representation, then, reflect a culture's understanding and conception of its environment. A representation echoes the criteria for ordering, the desire and beliefs, of the society that created it. In the last chapter, we saw that in the traditional representation of digital images (a 2-D array of pixels) was the desire for objectivity, and that intensity and adjacency were the criteria for ordering.

2.1.4 Embedded Knowledge

As much as representations reflect culture, it also reflect the *individual*'s beliefs, desires, inspirations and understanding. We have only to look into the works of artists, an impressive body of empirical evidence, to convince ourselves. Painting codes knowledge; it reflects the painter's knowledge of the subject, further reflecting his mental representation of reality.

The painter paints his knowledge.

Unintentional Coding

The encoding of knowledge is often done without the painter's (or, more generally, the author's) full intention, though commonly it is still too obvious for the reader to ignore.

In a letter Van Gogh [Gogh 67] wrote to Theo, circa mid-October 1888, he *describes* (codes, here in text) a painting of his bedroom, a place to "rest the head or rather the imagination".

> ...This time it's simply my bedroom, only color must do the job here...the walls pale lilac, the floor of a broken and faded red, the chairs and the bed chrome yellow, the pillows and the linen lemon-green and very pale, the cover blood-red, the washstand orange, the basin blue, the window green ...

Van Gogh, in this painting, codes his painting with color, having each color represent a particular cognitive object. He paints not what he sees, but what he knows and believes. True, Van Gogh is a genius, but it is hard for anyone not to be a genius in this respect—we have learned such a great natural power of cognition. The fact that there are words like "chair" and "window" reflects our mental representations of the world—we know such entities. Plato argues eloquently, in the *Republic*, and this is stronger than we intend to note, that in fact, the world is but a instance, and often a pure imitation, of the ideal world that is in the mind.

It is noteworthy that although Vincent lived alone, everything was painted in pairs, a pair of chairs, a pair of pillows etc. Although he

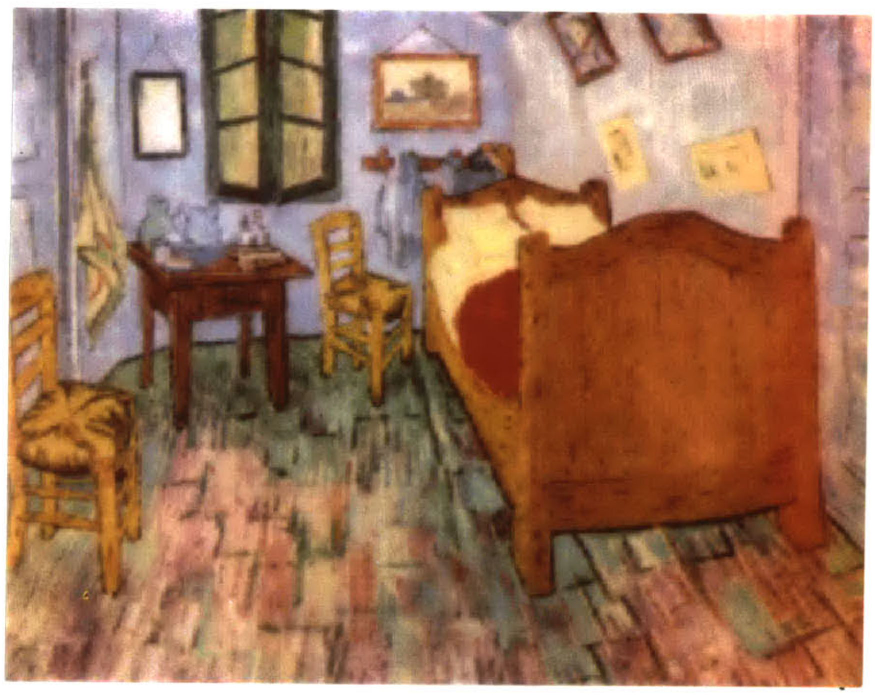


Figure 2.1: the bedroom, oil, arles, 1888

does not speak of it, it more than probably reflects his mind, exposing the intensity of his need for a companion. In fact, he did earlier write about other possible inclusions in the painting, "...there will be three subjects. Perhaps a nude woman, I have not decided, perhaps a cradle with a child; I do not know, but I shall take my time."

Intentional Coding

On other occasions, it is the author's explicit intent, sometimes even to the extent that it is the very philosophy of a school of thought, to explicitly include visual codes, to express its desires and beliefs.

The Cubists experimented with the idea of explicitly rendering not only what one saw but to also incorporate, within the rendering, what one knew of the depicted scene. In many paintings, an object is rendered, within a local canvas surface, from multiple points of views and moments of time.

Moholy-Nagy[Moholy-Nagy 65] writes:

> ...After the Cezanne's subtle attempts to create a new way of painting which would express such a vision, the avalanching power of the cubists brought forth a more accurate method. They showed the object, its elevation, plan and section on the flat canvas—as it would be seen from many viewpoints, in motion, revolving before the eyes of the spectator... [the] important feature [is the] visual force and emotional wealth of relationships, the constructive potential of the visual fundamentals.

The Cubists explicitly voiced that representation is inextricably linked with knowledge. The Surrealists tried to suspend knowledge by dismembering representation.

Knowledge is implicitly coded into representations. These "contents" reflect the desires and beliefs of a given culture. The study of how things are represented often gives us insights to how things are understood, and what is deemed important to a particular individual or society. The study of this coding is a basis for the study of culture. That representation inherently encodes interpretation, and that there is no hermeneutical neutrality in any representation is generally accepted.

In the next section we begin a survey of different levels and abstractions of representations of digital imagery.

2.2 2-D Surface for Image Representation

2.2.1 Cultural Image-Plane

In the formation of visual representation on the image-plane, the physical world is collapsed from its three-dimensional reality to its two-dimensional projection on a surface. If this "reduced" object(the image-plane) is to be re-read as construing the physical world, a reverse translation must be performed without the benefit of lost depth information. It is the absence of this depth information that causes image distortion as we view the image surface from different angles. In fact, there is but one view-angle from which the image-plane representation appears "correct". Then, to generally interpret this image plane correctly, which we seem to manage to do, we must "reconstruct" the *represented* from the plane-image and its visual distortions, appropriating our acquired perceptual experiences.[1]

2.2.2 In the Defense of a Surface

What are the qualities of an image plane that has made it such a successful and commonplace representation of the physical world?

In the earliest times, representation took the form of sculpture, a clumsy method, perceptually difficult to implement. Then came painting. The emergence of painting is not difficult to understand: people painted because they had walls! Simple as that.

Painting as representation of scenery met with success, and this lay

[1]It is possible that the human visual system has the innate capability to correctly interpret these planar images, and that this capability is triggered through experience. It is of interest to note that children begin to recognize such images only at age four.

largely in the fact that, because the representation is removed from the complete mapping of the represented, it acquired the strengths of the symbolic (as another largely successful representation, that of natural languages, is too) and thus *manipulable*, promoting deeper probing in the study of space; this is demonstrated in the succession of development of linear perspective in the Renaissance through Cubism and its further contortions at the present time.

It is not entirely surprizing that sculpture met with sudden death in the sixteenth century. Leonardo, writing in his Treatise on Painting [Leonardo 56], promoted painting, proclaiming it as supreme. He writes:

> If sculpture is lighted from below, it will seem monstrous and strange, but this does not happen with painting which carries all its elements within itself.... Applying myself no less to sculpture than to painting, and practicing both in the same degree, it seems to me with little change of bias I can judge which demands more of the mind, and which presents greater difficulty and is more perfect than the other.

Two-dimensional image-plane representation, in addition, is easy to implement and offers a large saving in bandwidth. This reduction in bandwidth allows the description of images to be produced either in reduced time or in higher fidelity. Together, these factors contribute to the success of this sort of representation.

However, discarding the third-dimension has its disadvantages as well. This discarded dimension could have been used to simplify various image-processing and tasks of understanding. Without it, large amounts of computation must be performed to extract otherwise seemingly obvious information. An example of such information is the contours of objects, which would have been a great help in the segmentation of plane images.

At this point, we would like to note that the following chapter, on *Methodology*, limits itself solely with cognitive associations of image surface areas of traditional image representations, avoiding three-dimensional models[2] and also sidestepping issues of three-dimensional reconstruction. Within such methods, as we just mentioned, are problems of segmentation computation. These issues will be skirted. The limitations are acknowledged.

However, we still allow ourselves to indulge, for it is our interest not to study the *represented* (the depicted scene) directly but to study it through its *representor*. Examining the vocabulary of a "limited"(less denotative) representation often gives more insights to relationships between thought processes and depictions of reality.

2.3 Abstract Surface for Image Representation

2.3.1 Memory Space for Intensities

The digital image is mapped to memory space.

Consider a monochromatic image. Using Cartesian coordinates, we represent intensity at point (x, y) by $B(x, y)$. We choose now a finite set of points (x, y) for representation, the brightness of each point numerically represented by a finite number of bits i.e. quantized (for example, using 8 bits per point, we can specify $256 (= 2^8)$ different intensity levels). This procedure is called Pulse Coded Modulation (PCM). The set of sampling points (x, y) is often a rectangular array, and we call the samples *pixels*. With the display of the pixels on the screen, we obtain a visual impression of the image.

One displays the pixels on the screen for the purpose of *presentation.*

[2]The problems mentioned might be transcended in the technology and art of holography

The "real image" lives in the abstract space of memory banks and registers. This is the classic and default convention of digital imagery.

2.3.2 Memory Space for $2\frac{1}{2}$-D Information

Other entities can be mapped to memory space.

This memory space for representation need not register image intensities alone. Bove[Bove 87] used the space to register depth as well.

For every position (x, y) in the image array, the albedo(intensity) rho and range data z is available. The aim is to code this information into the 2-D array of memory.

To begin, range data z are used to derive surface normals by differentiation. The slope of the surface in the x direction is called p, and the y slope is called q. By exploiting statistical coherence between p, q and rho, a visually reasonable 8-bit image can be obtained that includes all this data. The vector (p, q, rho) is quantized to find 256 representative vectors. The 8-bit image then references a 8-to-24 look-up-table whose "redness" codes p, "greenness" codes q and "blueness" codes rho.

This representation is based on "content" which is, instead of intensity to a fidelity of 256 values, (p, q and rho) with a combined fidelity of 256 values. This representation provides certain manipulability not otherwise possible. An example of this is the instant re-imaging given a new different light source. This is demonstrated here.

Assuming that light comes from infinity, the reflectance(and resulting intensity) of each point of the image-plane is simply

$$R(x, y) = \rho \hat{\mathbf{n}}(x, y) \cdot \hat{\mathbf{l}}$$

where $\hat{\mathbf{l}}$ is the light normal vector, and $\hat{\mathbf{n}}$ is the surface normal derivable from p and q by

$$\hat{\mathbf{n}} = \frac{(p, q, 1)}{\sqrt{1 + p^2 + q^2}}$$

Since there are only 256 distinct values of $\rho\hat{\mathbf{n}}$, the resulting image can be produced by just 256 multiplications.

2.3.3 The Frequency Space

An image can also be represented in frequency space.

But even in impressionistic descriptions, pixel intensity is definitely not the only property of concern. In particular, in certain analysis like edge detection, the rate of change of intensity from neighboring pixels is more useful. We note that the rate of change of intensity is within the domain of frequency descriptions. Thus, to enable ease of computation of edge detection, it is best to represent an image in frequency space.

To transform a set of data-points to its frequency-domain representation, we use the mathematics of the Fourier transform.

Suppose we have the image $B(x, y)$ with rectangular boundary i.e.

$$0 \leq x \leq X \text{ and } 0 \leq y \leq Y$$

Then we can describe the image completely as a sum of an infinite set of frequency components ϕ_{mn}, with weighting coefficients c_{mn}. [Netraveli 88]

$$B(x, y) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} c_{mn}\phi_{mn}(x, y)$$

$$\text{where } \phi_{mn}(x, y) = \tfrac{1}{\sqrt{XY}} \exp 2\pi\sqrt{-1}\left(\tfrac{nx}{X} + \tfrac{ny}{Y}\right)$$

The coefficients can be shown to be:

$$c_{mn} = \int_0^Y \int_0^X B(x,y)\phi'_{mn}(x,y)\,dxdy$$

($'$ denotes complex conjugate)

This set of coefficients forms the Fourier transform of the image, a representation in frequency space. It is a representation that makes frequency information explicit, and allows frequency related information to be easily retrievable. For example, it is trivial to deduce whether the image has large flat areas (equivalently, large low frequency coefficients), or if it is populated with fine detail (large high frequency coefficients). However, the spatial locations of these characteristics are not easily computable. In the frequency domain, images lose all spatial coherence in exchange for a gain in frequency coherence.

The frequency space can be used to represent images.

A problem with frequency space image representation is that it loses its spatial coherence. Visual cognitive cues are absent. Adelson [Adelson 84], develops an elegant representation that contains both frequency and intensity information. His representation, that he calls Laplacian pyramids, begins to break from the plane-image format into a multi-layered structure. His respresentation consists of a set of a pyramid of successively smaller images, with each image spatially coherent and containing a particular range of freuquency information. As in turns out, that a lot of simple cognitive processing that are computationally difficult in other image representations, are solved easily in this format. It is a more sophisticated reprensation and appears to be appropriate for purposes of image processing.

2.4 N-Dimensional Space for Image Representation

2.4.1 From Pixels to Predicates

The representation of imagery on a surface has its curses. It is perceived and manipulated as an *object*, an *analogon*, and such a perception blinds us from the fact that imagery is bound to cognition. Relationships within an image could not then be dealt with properly. The remedy to this problem is to abandon the *analogon* for a representations that are abstract and live in "cognitive-space", in n-dimensions.

We examine such ideas by making a quick ride through the motivations of digital imagery in its short history.

Pixels Found: Transferable

The desire for exact descriptions began with Mondriaan. His *de*-composition of images down to the use of rectangular figure blocks and of specific "primary" colors was, as can be argued, the beginning of the *pixelization* of imagery. This pixelization was a conceptual breakthrough and met, at least in retrospect, with great success. The process of pixelization was a desirable move as it lead to image representations that completely and exactly described images in a finite and small number of *signifiers*. These signifiers signified defined classes of entities, and the list of signifiers and their signifieds are known. This situation is analogous to quantized entry values of digital images and their referenced look-up-table.

Through the very finiteness of this type of representation, an image became transferable; an image could be transmitted in the sense that exact copies of it can be made and distributed. This has its lure. Through this development emerges the idea of a fundamental unit in images. In digital imagery, this unit has taken the form of the *pixel.*

A digital image is the sum of an array of pixels. The pixel is pure.

Pixels Reformed: Manipulable

At the same time, another aim in imaging was manipulability, with its beginnings in Renaissance Perspective and in Cubism. Only if manipulation were done on cognitive units would the results make any coherent visual sense. Thus it would be most appropriate to find representations that allow cognitive units to be readily represented. To represent cognitive units, we must code relationships between different parts of the image, for it is these relationships that delineate cognitive units in the image plane. *Relationships* are the key.

Predicating achieves this kind of representation. A *predicate* describes both the properties of the a local object(the pixel, for example) and well as its relationships with other objects, adjacent or otherwise. A predicate is, recursively, a grouping of pixels (or of other predicates) and the relationships between members of the group. The lowest "predicate" is the pixel and its properties.

The image-as-predicates allows coherence manipulations to be performed with much greater ease. A predicate, as a unit, can be transformed; relationships between predicates are trivially modified.

Predicates then replace pixels as ~~a~~ fundamental units or, more precisely, fundamental conceptual building blocks. Grouping of pixels is used here as a tool to signify that the relationships between pixels within the grouping hold~~s~~ more strongly than among others. A predicate could then be set to correspond to a cognitive object. Predicating offers a formal system that holds the issues of relationships central.

The image is a conglomeration of predicates. The pixel is united.

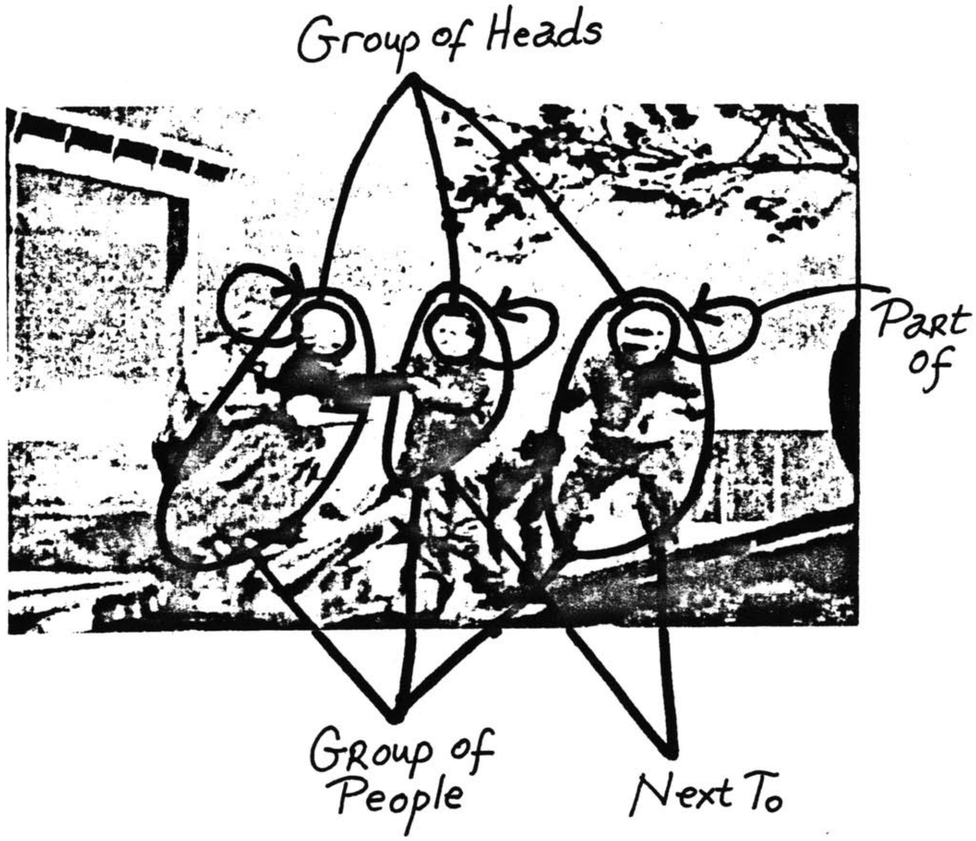


Figure 2.2: Predicates in Images

Pixels Lost: Relationships

It can be argued that the traditional representation of the array of pixels can also be read as predicates with the *adjacencies* as their relationships. It is even true that adjacencies in the image array are significant relationships, since adjacent pixels are, very often, constituents of the same cognitive entity.

However, adjacencies do not always necessarily indicate relevant relationships. The adjacent pixel often belongs to another cognitive entity. Adjacencies are also not sufficient; the same entity often encompass pixels that are not adjacent. Furthermore, compounded and higher-order relationships cannot be indicated in pixel array representations.

The pixel stands alone.

Pixels Regained: Knowledge

Do the individual pixels know they constitute a *dog*?

What use is a pixel by itself? It might as well be senseless, if it does not know that it belongs to some higher order. What use is a object if it does not belong to a larger order of things. An object out of context is not an object of meaning.

On the other hand, a pixel that knows what it is a part of, knows of other pixels. Knowledge spreads like wild fire. A representation that offers symmetrical readings (predicates know their constituents, and their constituents know what they belong to) would offer much strength and integrity to images; such a representation has a larger sense of completeness. Pixels gain character.

The representation with predicates makes exactly such dreams possible. Predicates are good. In coding a higher predicate of *dog-tree*, referring to the *tree* may guide us find the *dog*. And our *dog* knows that he is part of *dog-tree*.

Predicates add sense to otherwise nameless pixels.

There are examples of such coding in nature. A prime example is the DNA as the fundamental unit of living things. This unit of the DNA contains all the information of the organism that it is part of. It *knows* what it belongs to.

The pixel knows.

John Cage:

> Life without structure is unknown.
> Structure without life has no meaning.

2.5 Implicit Coding of Information

Another aspect of the implementation in the next chapter is implicit coding. We shall spend a few words here to speak about two simple precedents.

2.5.1 Explicit Information in Implicit Coding

A *representation* is a system for making explicit certain sorts of information or entities.

To formally complete a representation so as to enable its use and subsequent reading, it must be accompanied by a specification of how the system implements it. In other words, when we speak of *representation*, we speak of the *representor* and the *rules* that produce it from the *represented.*

Let us look closer at how *representations*, through their rules of construction, implicitly code certain formats of information thus making it trivial to explicitly extract precisely these types of information.

2.5.2 Representations of Natural Numbers

As an example, we can compare the Arabic and the binary numeral systems used to represent natural numbers. The Arabic representation consists of a string of symbols drawn from the set $(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)$. The method for constructing a particular integer n involves decomposing n into a sum of multiples of the powers of ten, then placing these multiples in a string, ordering them from the highest power on the left to the smallest power on the right. Thus, the number twenty-five would be decomposed to $2 \times 10^1 + 5 \times 10^0$, resulting in the string 25. The information that is most readily extracted from this representation (due, in this case, exactly to how it is constructed), is the various powers of ten of this number. The binary representation also consists of a string of systems drawn now from the set $(0, 1)$, and the construction rule is to decompose the particular integer n into a sum of powers of 2, ordering them in a fashion similar to the Arabic method. Twenty-five would be decomposed to $1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0$, resulting in the string 11001. It would not be a trivial task to extract the powers of 10 from this binary representation, though instead, the number's decomposition into powers of 2 is made explicit.

This purpose of this above example, to reiterate, is to demonstrate that the choice of representation affects directly the sort of information that can most easily be extracted. We notice too that, for a given representation, certain manipulations are favored. For example, multiplication by 2 in binary is trivial—just pad a zero onto the right. By induction, some complex processes could be trivial if they can be broken down to a series of compounded trivial manipulations.

The choice of representation is crucial. It determines what can be done with the represented.

2.5.3 Representations of Natural Scenery

The choice of different representation of visual material also helps to ease the extraction of desired information. A simplest example is that of maps. Different types of information are more easily extracted from highway maps than from landsat maps. We know which to choose if our goal is to drive our automobile through the terrain.

End Note

Having gone through a roller-coaster ride of the formats and issues of image representation, we are now ready to discuss the specific representation implemented. This is done in the next chapter.

Chapter 3

Methodology

Coding Content

In this chapter, we will discuss the implementation of the *content-based representation* of image sequences done in concert with this paper. The *content* encoded is image segmentations that spatially correspond to cognitive objects. This content is encoded not in a separate alpha-channel,[1] but, rather, implicitly within the representation; the content is embedded within the look-up-table.

The representation, specifically the look-up-table, is ordered based on this content: the representation is *content-based.*

Coding Context

In dealing with image *sequences*, we realize that the juxtaposition of images strongly affect their reading. *Context* makes *content*; the two

[1]red, green and blue channels represent their corresponding color components of an image; and alpha-channel is, generally, a fourth "channel" for other secondary information.

are inextricably interwoven. This is an especially pertinent issue in encoding digital image sequences, because here we are presented with the possibility of trancending the traditional *linear* ordering. Any image can be connected with any other.

We will code *context*, implementing it as *links* between images.

3.1 Why code content implicitly?

3.1.1 Why code content?

How do we *re*-present information given to us?

Given a certain amount of information, we make certain deductions to uncover what is of interest to us. We filter, select, and order this information to gain some understanding. These deductions can be anything from revealing insights to a simple re-ordering of data. From these deductions, knowledge accumulates. The *contents* of the information are these deductions.

In *re*presenting this information, it is the aforesaid deductions and their results that one would like to preserve, to record, to *re*present; it is the derived *knowledge* that we wish to code.

In re-presenting, it is ultimately the content we wish to convey. Thus, in representation of information, we want to code content.

3.1.2 Why code implicitly?

One common method of coding secondary information is to do so in a separate alpha channel. This coding would be *explicit* rather than *implicit*, in the sense that the information is contained outside and beside the image representation.

We have chosen an implicit method instead. The *content* that is being encoded is of central importance and should likewise be treated as the very core of representation. We then want representations which are based on content: representations that have content implicitly coded in them.

Implicit representations are like poetry—words living in themselves, while alpha-channel *explicit* representations are like annotated translations of poetry from a different language—parasitic and incomplete. Implicit representations are intuitively, conceptually stronger, more powerful and extensible.

There are certain advantages of implicit coding, including less overhead in decoding (as we shall see) and simpler file management (since they are identical with or without segmentation information being encoded). But the major reason for its choice is aesthetic: its simplicity and transparency gives it more elegance. Furthermore, it is in the direction of the, ultimately, model representation as we discuss in the next section.

3.2 The Model Representation

An image sequence is, at present, a sequence of frames, each of which is a mass of relatively unstructured pixels. An ideal representation would be one which is *highly structured based on cognition*, so that we may computationally extract what seems trivial to the human eye. What are the guidelines we should use to move towards our presently unstructured representation to one which is ideal? What are the limitations?

A model of such an ideal representation is imagery which is completely synthesized. Computer graphics is ideal. Computer graphics is a reflection of our cognition. How do we begin to transform image sequences toward descriptions of this sort?

We can begin by structuring frames. It is not difficult to computationally delimit a *shot* as a visually contiguous set of frames. The process is analogous for delimiting a long sequence into *scenes* and *acts*. *Shots* belong to *scenes* which in turn belong to *acts*. An hierarchical structure emerges. By treating *shots*, *scenes* and *acts* as scenic units, frames then become defunct. Frames are replaced by the duration of these *scenic* units.

Within a *shot*, are *characters* such as a dog or tree, and these *characters*, more often than not, populate other *shots* as well. These *characters* are object units. To correlate to computer graphics, the *object units* would be object descriptions while *scenic units* would be *motion* descriptions.

As much as objects in computer graphics are often composed of different sub-units, so too can *characters* be broken down into components. Conceptually, each pixel is eventually the smallest component. As with the synthesizing of computer graphics, each pixel is part of a *character* and its appearance is influenced by factors such as lighting and occlusion of other *characters*. With many frames as data, some of this information can, hypothetically, be deduced.

3.2.1 Limitations of Implementations

What are the limitations that keep us from the model representation?

In designing a representation, there is a set of tradeoffs between varying *constraints* and *desires*, having to realize the discrepancy between the ideal and the obtainable.

The desired representation, as mentioned, would be one that contains and incorporates innumerable cognitive principles. In representing an image, one constraint is that we aim for the retention of a format supplying a visual impression to the reader or, failing that, trivial conversion into such a format (computer graphics descriptions are an example). Yet another constraint involves the problem of cognitive knowledge and the difficulty of its formulation. As Ken Haase [Haase 90] describes

> "we must make certain that we neither separate objects which should be treated identically nor conflate objects which should be treated distinctly ... Consider representing a historical figure like Napolean; we could represent him by a single object, by a set of objects indicating Napolean at different periods of his life, by a set of roles (as general, as emperor, as revolutionary, as cultural hero) that he has variously filled, by different viewpoints on him by different authors ... we implicitly make important decisions about which differences matter and which do not. But there are no universal rules for determining this, but only rules which depend on individual goals ...".

The problems of knowledge representation are ultimately central in the study of content-based representation of digital image sequences.

3.3 Modus Operandi

For the purpose of demonstration[2] and for simplicity, the representation we implemented is a variation of the most traditional digital image representation, the *pulse coded modulation*(rectangular intensity array of pixels) representation, with the addition of encoding content. The *content* is image segmentation: image areas that correspond to cognitive objects.

To begin to encode segmentation, we need, first of all, to decide what constitutes an object in an image—what is a segmentation. It is not our concern here to compute automatic segmentation; these concerns belong to the field of image analysis and vision. Our aim is to encode ready-made segmentations. We will manually create our segmentations.

Automatic or manual, the delimitation of an image segmentation rests upon a *resemblance* of sorts between its constituent pixels. Resemblances between pixels such as adjacency, color, movement and gradient suggest that certain pixels belong to one cognitive object. Here, we will discuss only those guiding resemblances used in our implementation.

The idea of *resemblances* as a guide for classification and the formulation of knowledge is not novel. In fact, it was the central *episteme* in sixteenth century Europe, and by some coincidence there are large similarities between their veins of approach and the techniques pursued in our implementation. We choose then to discuss our techniques in parallel with facets of this *episteme.* There are no claims that this manner is of any significance, but formatting discussions within such confines might enable examination of subtleties otherwise unseen.

[2]just say no!

3.3.1 Resemblances

Up to the end of sixteenth century, resemblance played a decidedly constructive role in Western knowledge, to the extent that not only was it the tool for representing knowledge, but it also made possible the knowledge of things concrete and abstract. Foucault[Foucault 74] notes:

> Resemblance made possible knowledge of things visible and invisible and controlled the art of representing them. The universe was folded upon itself: the earth echoing the sky, faces seeing themselves reflected in the stars, and plants holding within their stems the secrets that were of use to man.

The four similitudes are:

- *Convenientia*—Resemblance by adjacency.
- *Aemulatio*—Reflection at a distance.
- *Analogy*—Convenientia merged with Aemulatio.
- *Sympathy*—Free State Linking of knowledge.

Each of these similitudes parallels the techniques used in our image sequence representation. *Convenientia* and *Aemulatio* are used for image segmentation through spatial and color coherence respectively. *Analogy* conceptually echoes the deduction that near-similar segmentations of single frames are really just one cognitive entity that lives across frames. Finally *sympathy* echoes the links between image sequences. The rest of this chapter discusses each of these items.

3.4 Convenientia

Convenientia denotes adjacency—a resemblance of place.

Sixteenth century thinkers considered adjacency as only a sign and visible effect of the real hidden and stronger relationship. It is no accident that things are adjacent; there is a good plan for it. Nature placed things next to one another because these things have a deeper of level of similarity. Thus it is no accident that one part of a finger is adjacent to another part of a finger, and that one finger is adjacent to yet another.

Convenientia is the first principle for ordering of the visible.

3.4.1 Spatial Masking for Segmentation Synthesis

Adjacencies offer cognitive guidance in identifying objects in Nature, and this is no different in the image array.

We will use spatial coherence as a guide to manually create segmentation. We segment an image spatially into parts that roughly correspond to cognitive objects. One part may correspond to a dog, another part a tree, and the third the background. This is segmentating with spatial cues.

This is a very simple procedure. A paint program is used to manually synthesise the segmentation by painting spatial areas for each cognitive object. A 1-bit mask would be able to denote 2 segmentations (background and foreground). A n-bit mask can denote 2^n segmentations.

This reordering of the image into spatial areas must now be *encoded* into the description of the image. In fact, these spatial areas (the content) must be used as the basis of the representation of the image.

3.4.2 Encoding

We will demonstrate how to encode an 8bpp image A with 2 segmentations written in the bit-plane M.

The *explicit* method would be to simply have the bit-plane M as the alpha channel. This would need an extra bit per pixel bandwidth, which is a large amount considering some image compression techniques can compress 8bpp images down to 0.2bpp.

Our *implicit* method will require no extra bandwidth and will cause only the most minimal image degradation. We will encode the segmentation within the look-up-table.

Look-Up-Table Concatenation

The technique is called look-up-table concatenation.

The general idea is that we wish to rearrange the image such that each segmentation of the image reference only one contiguous section of the look-up-table. The algorithm below and figure 3.1 demonstrate how this is achieved.

1. Use stencil to produce 'background' and 'foreground' images, A_1 and A_2. For each pixel p,

$$if\ M(p) = 1,\ then\ A_1(p) = A(p),\ else\ A_1(p) = 0$$

$$if\ M(p) = 0,\ then\ A_2(p) = A(p),\ else\ A_2(p) = 0$$

2. Compress[3] both look-up-tables linearly (or some fancier compression method) such that

$$length(lut(A_1)) + length(lut(A_2)) \leq 256$$

3. recombine A_1 and A_2 by collaging to produce reconstructed image R, offsetting one pixel values of the second segmentation by the length of look-up-table of the first.For each pixel p

$$if\ A_1(p) \neq 0,\ then\ R(p) = A_1(p)$$

$$else\ R(p) = A_2(p) + length(lut(A_1)$$

4. concatenate $lut(A_1)$ and $lut(A_2)$ producing $lut(R)$

5. annotate in the header of resultant image R, with $offset$ which is equal to $length(lut(A_1))$

3.4.3 Decoding

The decoding could not possibly be simpler. The typical situation is that we point to a particular pixel in image R, and ask which segmentation it belongs to.

Here is the algorithm used.

1. read once, in the header of R, the $offset$

2. given pixel p,
if $R(p) < offset$ then it belongs to segmentation 1,
else it belongs to segmentation 2.

[3]8 bpp images cannot refer to entries larger than 256.

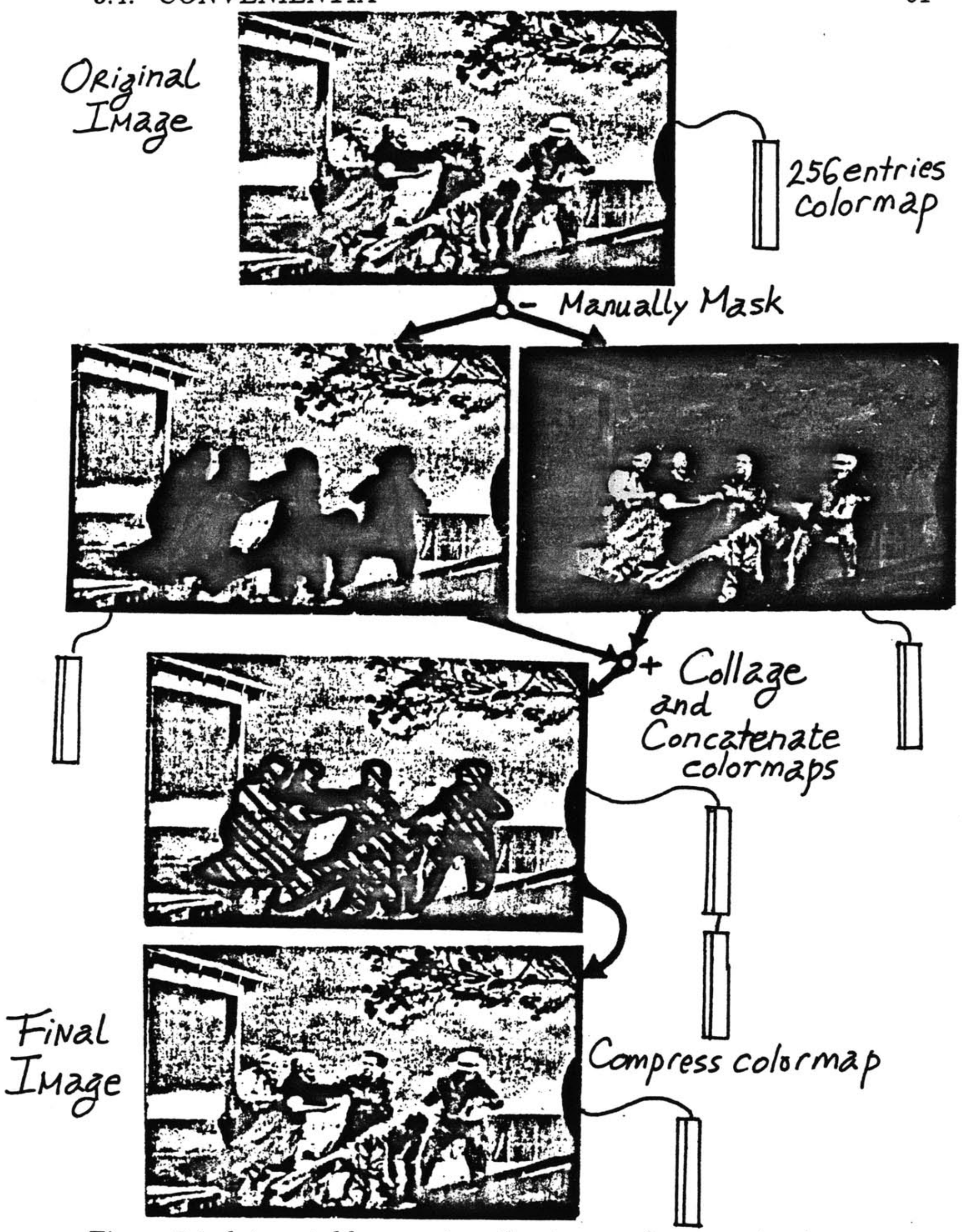


Figure 3.1: lut-up-table concatenation to encode *convenientia* segmentation

3.5 Aemulatio

Aemulatio denotes similarities between items at a distance.

Freed from the bounds of adjacencies, *aemulatio* is a form of *convenientia*, broken from the grips of spatial laws to function at a distance. It is similitude transcending space, amounting to reflections and conversations between these 'disjoint' objects, entirely without motion. Something else in the commonality between disparate (distanced) objects that communicates communality. It is really *convenientia* on a different dimension—on a non-spatial dimension—as though it is their references that are in adjacency. In this sense, *Aemulatio* is a second-order ordering of the visible.

> The human face, from afar, emulates the sky, and just as man's intellect is an imperfect reflection of God's wisdom, so his two eyes, with their limited brightness, are a reflection of the vast illumination spread across the sky by sun and moon; the mouth is Venus, since it gives passage to kisses and words of love; the nose provides an image in miniature of Jove's sceptre and Mercury's staff.[U.Aldrovandi, Monstrorum historia, (Bononiae, 1647 p.3)]

3.5.1 Colormap Masking for Segmentation Synthesis

Similarities between pixels need not derive from spatial adjacency, but could arise from adjacency in intensity, in local gradients, in hue, in texture etc. In the case of hue-similarity, the pixels reference adjacent entries in the look-up-table: it is an *aemulatio*. This similarity can guide extraction of cognitive objects from the image array. Many cognitive objects do encompass only a limited range of hue—a red shoe, for example, or even skin tones, and thus we can use hue for classi-

Figure 3.2: Original image displayed with its colormap.

fication and perform image segmentations. In fact, there are certain types of segmentations that are more conveniently created using this method than the previous one. An example is a flock of birds, where the segmentation is not spatially localized.

We synthesize a segmentation by tagging the entries of the colormap[4] that are referenced by a cognitive object. We create a colormap "mask" for 2 segmentations, using this algorithm. We begin with the original colormap M, a temporary copy C, and end up with a list of tagged entries L.

1. Make a copy of colormap M, calling it C.
2. Use mouse to pick a pixel on the part of image that is within the desired segmentation. Note pixel value.
3. For the purpose of a visual check: change corresponding entry in colormap C to *highlighted*; loading this colormap will display all pixels referencing, highlighted. If 'wrong' pixel was picked, resulting in wrong pixels being highlighted, undo this step and goto Step 2.
4. Add entry to a list of tagged color-map entries L.
5. Repeat from Step 2 until satisfied.

3.5.2 Encoding

We need now to encode the segmentation in our representation. Since the segmentation information is completely contained in knowing the the tagged entries L, all we need to do to encode the segmentation is to create a new colormap R that groups together the tagged entries. This is a simple rearranging/reordering of M, by having all tagged entries in the first section of R, and the rest in the second section.

[4] A colormap is a look-up-table of colors referenced.

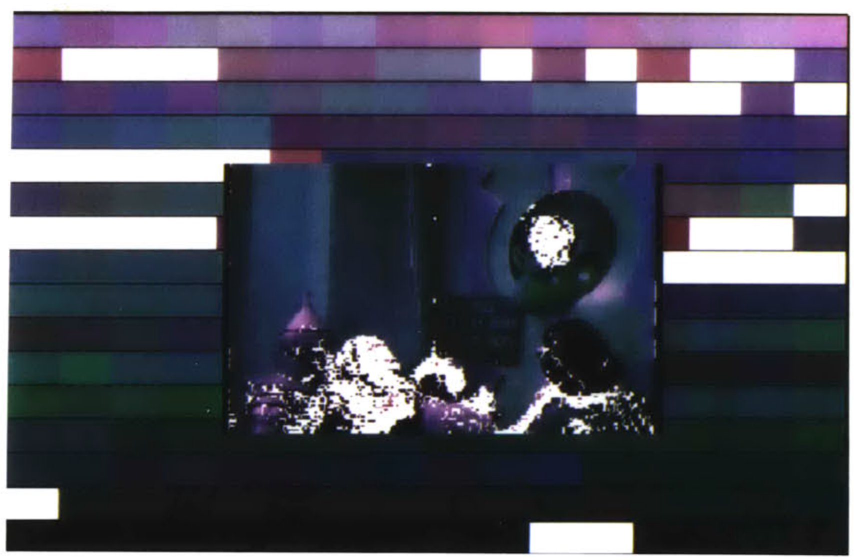


Figure 3.3: Image Segmentation displayed with tagged colormap.

1. Copy first, all tagged entries L into a new colormap R.
2. Concatenate all other entries of M into R.

Rearranging the colormap requires corresponding adjustments in the pixel array. Since an entry has moved its position in the table, all pixels referencing this entry have to reference this new position.

1. If pixel value is greater than first entry in list of tagged L, increment pixel value by one.
2. If pixel value is greater than next entry in list of tagged L, increment pixel value by one.
3. Repeat from Step 2 until pixel value is smaller than value of next entry in list of tagged L.
4. Repeat from Step 1 for all pixels in image array.

In this case of encoding 2 segmentations, there is absolutely no extra bandwidth needed, nor is there any information lost. A simple reordering incorporates this new information, without sacrifice of anything. We have gotten something from nothing!

For more than 2 Segmentations

In the case of encoding more than 2 segmentations in the image array and colormap, we encounter the problem where more than one segmentation might lay claim to a colormap entry. And because distinct sections of the colormap represent distinct segmentation, we must, in the case of the foresaid situation, have *that* particular entry in both sections: we have to replicate entries.

Surely there is no problem with replication. Except for one: we are limited, in images of 8 bits/pixel, that there is a ceiling of $256 (= 2^8)$

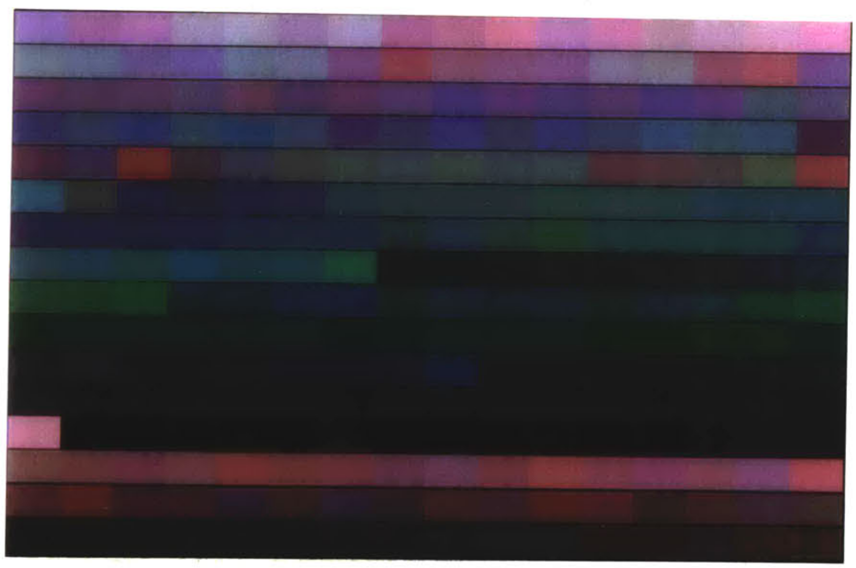


Figure 3.4: Rearranged colormap with two segmentation. The numbers in the final image is changed to reference the correct entries, but the image is visually identical with the original.

entries in the colormap. Thus a consequence is that each time we add an extra segmentation(with its accompanying extra entries), we need to squeeze the colormap so as to keep the total number of entries within the allowed limit.

Another simpler method is to begin with a *virgin* image that has a compressed colormap of an arbitrary 150 entries. Each time, a new segmentation is added, extra entries are concatenated, and when it reaches the maximum allowable number, no more segmentations are allowed.

We see that, in encoding more than two segmentations, no extra bandwidth is required but a small amount of hue resolution is sacrificed.

3.5.3 Decoding

In the general schema of incorporating knowledge in descriptions, order is increased. There was an intent of ordering according to certain principles, and the results of the scheme reflect this process of ordering. Ease of reading is enhanced because of this ordering.

Bateson[Bateson 72] writes:

> Order is seen as a matter of sorting and dividing. But the essential notion in all sorting is that some difference shall cause some other difference at a later time. If we are sorting black balls from white balls, or large balls from small balls, a difference among the balls is to be followed by a difference in their location—balls of one class to one sack and balls of another class to another. ...[a perceiving Entity must be] invoked to perform this function of creating an otherwise improbable order.

This reordering of data based on *content* makes the extraction of this content computationally very easy. Here, the decoding is identical to the one done for the *convenientia* technique.

3.5.4 Implicit Constraints

As much as *implicit* coding has its merits, there are associated problems, some of which are actually absent in *explicit* methods. Since the *content* is embedded into the ordering of the image, any processing that manipulates and disturbs this ordering would in effect eliminate what is stored. For example, the segmentation information would be lost as simple process as a low-pass filtering.

However we must contend that this problem is only indirectly due to implicit coding, but a problem that arises from the nature of colormaps. Except for grayscale colormaps, filtering on 8bpp images does not generally work. Filtering works only when the values of the image array are proportionl to image intensity, and this is the case for 24bpp or 8bpp grayscale images. The fact that we have decided to encode segmentation information into non-grayscale 8bpp colormaps, constraints us the the limitations of the latter.

Thus before choosing where to embed our information, we should know for what eventual purpose the image is. We put our carrots either in the ground or in the rabbit's cage depending on what we want the carrot to do.

3.5.5 Texture Resemblance as Segmentation Synthesis

Let us now embed our information in the look-up-table of vector quantization [Gray 84] instead of a colormap. We choose this because we know that we would like to eventually compress the image, and vector quantization allows us just that.

The *aemulatio* resemblance in this case is of texture. We start with a, say, 300×400 pixels 8pp color quantized image A. Our algorithm is as follows:

1. Divide the image A into 2×2 blocks to form a quarter sized array V_o.

2. Make a histogram of the (2×2)-dimensional vectors of array V_o. There should less than the maximum $\frac{300 \times 400}{2 \times 2}$ vectors.

3. Guided by the histogram, select the best representative 1000 vectors. These vectors form the codebook, B_o.

4. Run through V_o and replace the each vector with the best approximate vector from the codebook B_o.

5. Now we are ready to embed segmentation information. Reconstruct vector-quantized image R_o, and display. Click on parts of image of intended segmentation. For each click, find corresponding entry in V_o, and this entry's entry in B_o. Tag those entries of B_o.

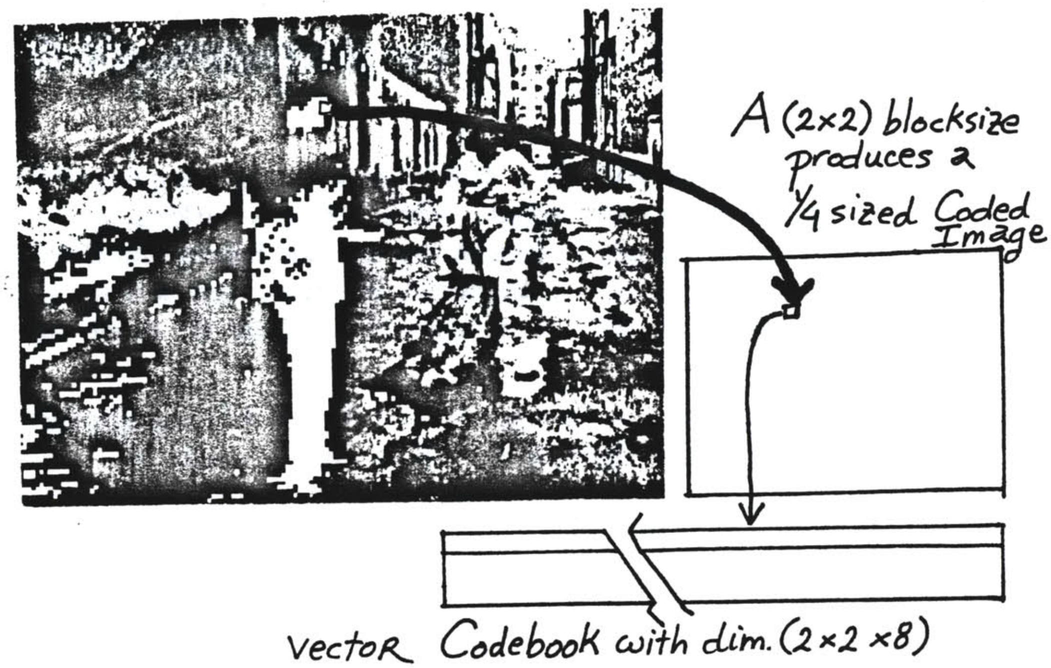
6. Rearrange B_o to create B_r by grouping all tagged entries to the top of the codebook.

7. Recode V_o to form V_r so that they point to the correct entry of the rearranged codebook B_r.

8. Reconstruct vector-quantized segmentation embedded image V_r, which should look visually identical to V_o.

   The decoding is similar with that of *aemulatio* hue-resemblance image.

   As noted in the last subsection about constraints, this image, embedded with segmentation information, can do all that could have been done to vector-quantized images.



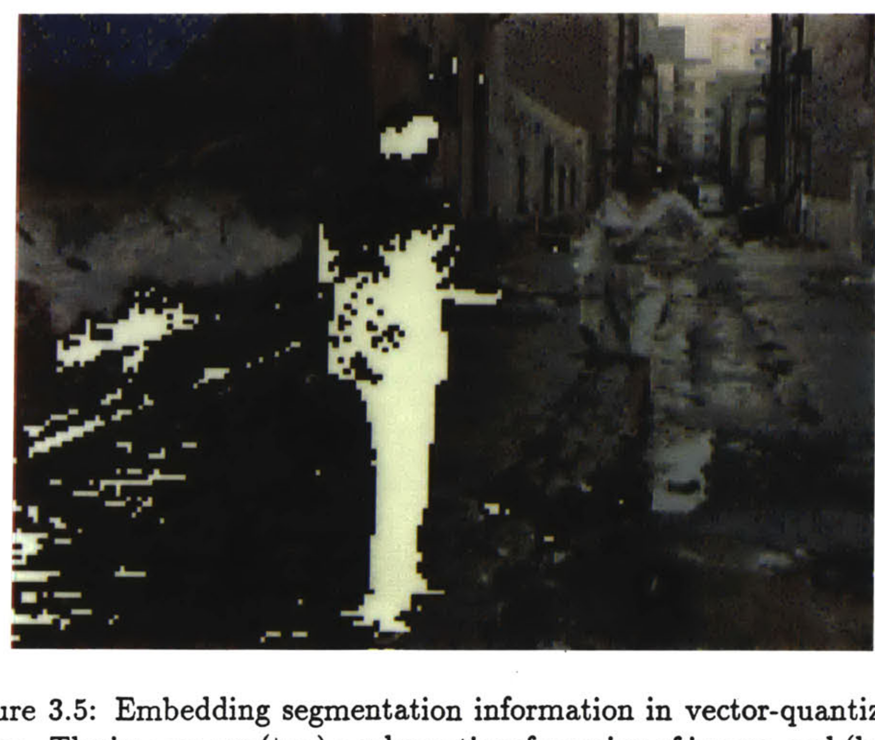


Figure 3.5: Embedding segmentation information in vector-quantized image. The images are (top) a schematic referencing of image, and (bottom) an image with segmentation produced by textural resemblance.

3.5.6 Application: Segmentation-based Resolution

It is often desirable for a certain part of a moving sequence to be rendered in more detail. When a dog is barking up a tree, we might want only to look at the tree and ignore that silly dog. When a cow jumps over the moon, we might want to stare at the patterned flying cow - there is not much about the moon we can describe anyway.

It is often true that the transmission channel for images is narrow. We cannot pipe through everything that we want. But what do we really want to pipe through anyway? Perhaps all of tree and just a little of mad dog.

3.6 Analogy

Analogy is *convenientia* and *aemulatio* superimposed.

We have seen how *convenientia* and *aemulatio* have been guiding rules in synthesizing image segmentations within the still frame. *Convenientia* is a first-order spatial adjacency, while *aemulatio* is a second-order adjacency in the sense that the referenced was adjacent. *Analogy* will help us transcend the frame.

In an *analogy*, the adjacency is more global. An *analogy* between two things is a resemblance in structure, a resemblance in relations. For example, there is an analogy between the human skeleton and that of birds:

> the pinion called the appendix which is in proportion to the wing and in the same place as the thumb on the hand; the extremity of the pinion which is like the fingers in us ...; the bone given as legs to the bird corresponding to our heel; just as we have four toes on our feet, so the birds have four fingers of which the one behind is proportionate to the big toe in us. [P.Belon, Histoire de la nautre des oiseaux. Paris 1555 p.37]

There is a *convenientia* resemblance of one point in the human skeleton and a corresponding point in the bird's skeleton. There is also an *aemulatio* resemblance in the structure of the two skeletons.

3.6.1 Analogies-derived Units

Through *convenientia* and *aemulatio*, we have created image segmentations for a still frame. On successive frames, we will also likely obtain similar segmentations. With a dog in one frame and a dog in the next frame slightly displaced, we make a correlation and claim they are not to be two objects, but one and the same.[5] Encouraged by a *convenientia* in time(frames) we make an *analogy* between the dogs in adjacent frames. There is but one dog and it is unleashed from the bounds of the frame. We call the dog, this entity that lives across frames, a *character.*

The frame is superceded in this representation.

These analogies can be taken on further, producing *analogies* on a higher order. By viewing the imagery and further relations and groupings of events and objects, parts of the aforementioned *model representation* can be attained.

3.7 Sympathies

The final *resemblance* is *sympathy*, a similitude that is formed without any set prior rules prescribed. *Sympathy* "plays through the depths of the universe in a free state"[Foucault 74]. It could be formed through a simple contact as 'mourning roses that have been used at obsequies' which, simply from their former adjacency with death, will render all persons who smell them 'sad and moribund'. [G.Porta, Magie naturelle (Rouen 1650 p.72)]. But *sympathy* is not restricted to any form. "No path has been determined in advance, no distance laid down, no links prescribed." And it is perhaps of this free form that elevates this instance of the *Same* to such a strong and insistent level—it allows the reader of the Universe to prescribe links with minimal effort, to allow him to assimilate and integrate his knowledge.

3.7.1 Associations and Links

We code *context* as *links* between associated images. These *links* are *sympathy* similitudes.

[5] Much has been written about identity and spatio-temporal continuity[Higgins 67].

Montage, in linear films, has been the primary method in prescribing links between image sequences to create specific readings. Eisenstein has studied, with great scrutiny, the effects and fundamentals of editing, demonstrating it, with spectacular insight, in the film *Potemkin.*

Laser discs offered film-makers a broader set of options upon which sequences are presented. No longer was it necessary to limit oneself to linear screening—there was the added ability to produce movies with multiple (possibly intertwining, repeated etc.) paths, and the choice of journeys could be chosen dynamically by the viewer by interaction, or set statically by the maker. In the Interactive Cinema (*ic*) Group in the Media Lab at M.I.T., work has been done to enable the user himself, using presented *tools*, to create his own paths while he explores the material, affecting its subsequent readings[Davenport 89]. These paths are laid based on *sympathies* one expresses.

While the *links*[Davenport 89, Drexler 87] in analog laser discs done at the *ic* Group is a frame-to-frame link, the *links* here of digital movies are more general. First of all, an advantage to digital movies is that we need not be limited to accessing particular frames, but may access items within the frame—the characters, the individual pixel etc. The default *link* implemented in our program links a character with some other entity—a data-base, a visual effect, or another movie sequence. *Tools* have been implemented to allow the reader of the movie to create *links.*

3.8 Signatures

How does the reader know that certain content has been encoded?

Knowledge, as it is being extracted from experiences and discoveries, and re-encoded back into our descriptions of the world, would be left unread and unnoticed, and thus unutilized, if we were not to make a mark that it be recognized. As we develop our understanding of the world around us, so then are our libraries filled with manuscripts. These manuscripts are descriptions of the world, serving as signposts of our journeys, and enables us, as readers, to re-read what has been coded.

> We might make our way through all this marvelous teeming abundance of resemblances without even suspecting that it has long been prepared by the order of the world. In order that we may know that aconite will cure our eye disease, or that ground walnut mixed with spirits of wind will ease a headache, there must of course some mark that will make us aware of these things... Would we ever know that there is a relation of twinship or rivalry between a man and his planet, if there were no sign upon his body or among the wrinkles of his face that he is an emulator of Mars or akin to Saturn? [Foucault 74]

We need to label the contents, and allow the reader to read these labels if he needs to.

3.8.1 Names as Signposts for Knowing

In encoding segmentations implicitly into digital images, the problem arises whereby there are no visible signs to indicate what has been encoded. Whether or not a dog has been encoded in the image is not known by the human reader. Names to each segmentation, acting as signposts of knowing, must be hidden someplace, and we do it here in image descriptor files that can be readily read.

> It is not God's will that what he creates for man's benefit and what he has given us should remain hidden... And even though he has hidden certain things, he has allowed nothing to remain without exterior and visible signs in the form of special marks—just as a man who has buried a hoard of treasure marks the spot that he

> may find it again. [Paracelsus, Die 9 Bucher der Natura Rerum(Works, ed. Suhdorff, vol.IX, p.393)]

In the work done in the *ic* group, MICONs(motion icons) are presented to signify to the reader the presence of encoded knowledge. A MICON tells the reader that the machinery has delimited and knows of a particular object or sequence, prompting him to *select* them if it is his will to do so. This is effective and visually appealing. An odd aspect of MICONs lies in the fact that they are often but miniatures of what they identify. To select a visual object on the screen, one must click, rather than on the object directly on an icon that signifies it.

The advantage of our digital counterpart is, because we can access intra-frame units, we no longer need MICONs. We interact directly with the image itself.

End note

Thus we have encoded *content* and *context* within the image representation.

We have written the image, and in the next chapter focus on the issues of reading.

Chapter 4

The interactive movie

The last three chapters focused on representations—how images are described and written. In this chapter, we turn our attention to how these represented images are subsequently read. This is most important since, ultimately, images are represented to be read. In fact, it was precisely because of concerns about the eventual reading that motivated the choice for content-based representations.

We will discuss in this chapter, a particular kind of reading—the reading through and of an interactive movie. Here, the reader interacts virtually with the writer through the movie. We will examine the nature of reading an interactive movie to discover some of its associated problems and their sources. We will proceed to formalize an interactive movie as a *text* and construct a computational model of the movie, its reader, and the communication channel. With such a model intact, we can suggest certain approaches to solve those problems.

4.1 Of links, scripts and the interactive movie

Towards the end of the previous chapter of **Methodology**, we discussed the implementation of links as *context* operatives within a content-based representation of the moving image. Computationally, these *links* are simple pointers that may be set, assigned to specific functions and programmably "activated".

These *links* can connect otherwise disparate *scenes* and *players* of the representation, to each other, to exterior textual and visual databases, or to operative functions. These *links* can specifically be used as *jump cuts* while the movie is being *played.* In such a use, a *link* functions as a montage[1] link.

A set of links, in this function, is a set of montage points. Such a set plays and replaces the traditional role of the linear script, but is more versatile: it allows for a *graph*-like[2] script, plus any programmable functionalities such as access to databases or triggering of visual processing effects. Thus *scripts* are incorporated within our content-based movie representation.

These *montage links* of our script can be either *hard-wired* or *soft-wired*, that is to say, the link either is always active or awaits activation—remaining dormant, *consummated* depending on some state. This *state* that determine the consummation of a link, the transformation of pointers as links into montage links, can be configured to be affected by the reader. Through this, the reader gains some control of the path the movie takes. The reader creates an instance of a script. We begin then to have an interactive movie.

Thus begins a discourse between the reader and the movie machin-

[1]montage a la Eisenstein: a structured edit that builds meaning through the metaphoric associations the reader is encouraged to make.

[2]web-like; with nodes and edges; binary tree is a type of graph.

ery, which is the writer in virtual presence. This interactivity between human and machine (the *virtual writer*) happens through the language of a movie. We must examine the nature of this interactivity, its implications, its effects, and its weaknesses.

4.2 Impact of the interactive

The nature of interactivity affects how an image is read.

The layout structure of information and its access in interactive media differs from that of non-interactive media. In the former, in all its varying degrees of interactivity—from occasional prompting to continuous control—the reader is guided through the material, and prompted to search and select material of their interest. This is not so for the latter.

In general, it is most likely that the impact of interactive media[3] (the apparent amount of understood and retained material), is larger than that of non-interactive media such as linear video. This has been demonstrated in various applications, from foreign language study discs to (re)educational discs for professionals.

It has been argued that different media tend to use and could potentially cultivate different thinking skills[Olson 74], that instructional media function as activators and potential cultivators of cognitive styles and strengths. Interactive media is a poignant candidate in these realms. Although some exploration is being done in this direction[Gagnon 86], this relation between interactivity[4] and learning methods is still largely unexamined.

[3] The observations are of *analog* laser disc interactive movies, but these observations hold true for digital or analog. For the rest of the chapter there will be no distinction between them, unless specifically stated.

[4] Constructivist approaches to knowledge examined by the Epistemology and Learning Group at the Media Lab works best. Constructivist, learning through building, encompasses interactivity.

4.3 Open and Closed Texts

An interactive movie can be considered and studied as a text.[5]

Eco[Eco 79] distinguishes between two *texts*, the open and the closed. An open text is one whose interpretation is restricted and the *Model Reader*[6] is highly defined and selected. An open text can be read only by readers with sufficient knowledge. A closed text, on the other hand, is one where the *Model Reader* is a general reader. It is a text that is "frequently interpreted against the background of codes different from those intended by the author."

We see here parallels between these two types of *texts* and the two types of interactive movies made to date. On the one hand, instructional interactive movies like "How to Brush your Teeth" are open texts because it allows for little interpretation, and the intentions of different readers are not divergent. That its intended audience is consciously selected, these texts generally perform well in their often well-defined purposes.

General entertainment interactive movies, on the other hand, are closed texts, for there can be an immoderate range of interpretations amongst its varied audience. That its intended audience is wide, and that the goals are less determined, it is then not entirely surprising that these movies tend to be less successful at their goals. There are many reasons why such movies fail.

In the rest of the chapter, *we will be focussing on movies that are closed texts*, to examine the reasons why they tend to be less than entertaining.

---

[5] *Texts* is a "network of different messages depending on different codes and working on different levels of signification".

[6] A *Model Reader* is a model of the possible reader.

4.4 The role of the reader

How is a movie read in interactive systems?[7]

The standard communication model constructed by information theorists is tripartite—the *sender*, the *message* and the *addressee* [Eco 76], in which the message is decoded in reference to a code that is shared between sender and addressee. However, this model does not describe the actual action of the discourse, especially in the case when the effective *message* is cooperatively generated by the reader[8], of which the interactive movie is an example *par excellence.*

As the reader forays amongst the many *paths* through abundant material, he wanders, skims, trips, and occasionally discovers personally engaging segments, creating in the process the construction of a script and an accompanying message. The reader searches for material of interest, for material of relevance. It is, in some senses, a quest for identification with the material. To some extent, a reader reads to educate and inform himself of particular issues. To a larger extent, a reader reads to reinforce his ideas about the particular matters[9], to buttress it with facts and stylistic eloquence. He identifies with parts of the material, and in his identification he indulges in knowing that other readers, in their reading, are exposed to and informed of "him" (the image of things with which he identifies). He becomes *famous.*

In reading, *the reader searches for himself.*

The failure of such a quest results in disinterest and a lack of abandon, either because the material is irrelevant, or the structure of presentation inaccessible.

The reader has to work to mentally transform the original literal

[7]From this point, all references will be to movies as closed texts, unless stated otherwise.

[8]*reader* and *addressee* used interchangebly.

[9]True, often in a reading he modifies or even changes his opinions and insights.

messages to what he finds desirable, or lack of that, what he believes the writer intends him to understand. Allegorically, this is demonstrated by Michel Tort, in his book Intelligence Quotient.

> What is going to determine the answer to the question is not the question as such in the form in which it was posed, it is also the idea that the interrogated subject forms; about the most appropriate tactic to adopt in function of the concept he has formed about the expectations of the interrogator.

What happens when a black man watches a soap opera about white middle-class suburbia? There might be no decoding to form messages of relevance to him, no possible consistent transformation of the literal message. When this situation arises, the reading ceases.

4.5 The responsibilities of the writer

The interactive movie as a *closed text* often is interpreted against a set of codes different than the one intended because the *actual* reader is different from the model reader the writer had in mind. The writer did not take into account such a situation. It might be acceptable that this results in aberrant decoding. But worse yet, the movie might become irrelevant to the reader who in turn ceases to read. The *closed text* unintentionally becomes an *open text.* The reader becomes dissinterested and disappointed.

The reader is uninterested—thought he is not aware of the reason for his lack of interest—because he has been *ignored*[10] by the writer, who although might have intended the *text* to be for the general public, had failed to include this particular reader. It is particularly annoying that

[10]In the absence of a gender-neutral noun in English, "he" is used to mean *she* and *he.*

the reader senses, from the type of text, that the writer had intended to speak to all. The reader feels not only excluded—he is *rejected.*

This observation is of course true for all media, but is of particular issue in interactive media, for it is these media which lay its very foundations and claims on the *active participation and incorporation* of the reader.

Inspite of the presumptions in interactivity about exchanges between the two parties across the interface, the reader's input is often not sufficiently acknowledged or incorporated into the system.

Baudrillard[Baudrillard 83] writes:

> The role of the message is no longer information but testing and polling, and finally control, in the sense that all your answers are already inscribed in the "role", on the anticipated registers of the code. Montage and codification demand, in effect, that the viewer construe and decode by observing the same procedure whereby the work was assembled. The reading of the message is then only a perpetual examination of the code.

Eco[Eco 79]:

> They apparently aim at pulling the reader along a predetermined path, carefully displaying their effects so as to arouse pity or fear, excitement or depression at the due place and at the right moment. Every step of the "story" elicits just the expectation that its further course will satisfy. They seem to be structured according to an inflexible project. Unfortunately, the only one not to have been 'inflexibly' planned is the reader.

While reading an interactive movie, we must select the material we are presented with. But then conversely, the readers are selected by the writer. In today's software industry, the code is written by a demographically small range of people, and thus correspondingly the resultant software is also for a very particular sample of audience.[11] Following the tradition of cinematography, the socially disenfranchised are ignored. There is often nothing in the material that is specifically of interest to minorities, and thus no reason for interaction. Interactive movies, even in their conscious pretense of reader incorporation, do not transcend the practices of an segregated society.

The writer, possibly unaware, is often dangerously ethnocentric. There needs to be some incorporation of the study of sociology, of sociolinguistics, in the design of interactive systems and interfaces.

The writer has responsibilities.

4.6 Computational design

In this section, we will formulate computational designs that address and possibly rectify the problems of interactive movies discussed in the last section. What computational structures and strategies are needed to guide the programming of the movie machinery so as to produce interactions with substantial communication between the machinery and its readers. Given a mass of material, ethnographically-biased or not, what are the approaches that best incorporate the reader?

4.6.1 Exchanges between graphs

To begin to appreciate the source of some of the problems of the interactive movie, we must formalize the structure of interactivity and study its components.

[11]It has been shown that, in general, access to high-technology by minorities is limited.

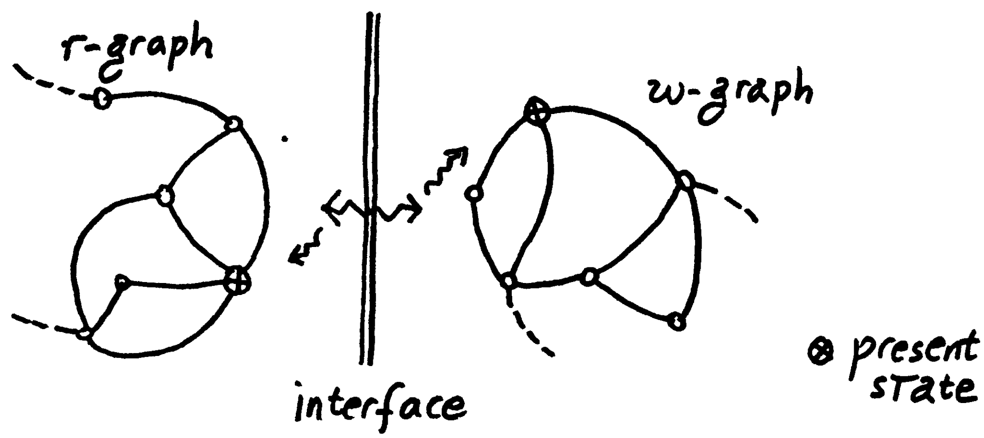


Figure 4.1: w-graph and r-graph

An interaction happens through an *interface*. An interaction happens between two (or more) parties. In our discussion on interactive movies, the two parties are the reader and the *virtual writer* represented by the movie machinery. The movie itself is the *language*, the format of discourse.

The two sides of the interface can be computationally modelled as two *graphs* with a communication path between them. One graph represents the states of the movie machinery, the states of the virtual writer. The other graph represents the *reader*. We shall henceforth call them the *w-graph* and the *r-graph* respectively.

In programming of an interactive movie thus far, much attention is focussed on the *w-graph*—the structure of actions and reactions of the computer in the exchange. Little attention has been given to the modelling of the *r-graph*—the anthropological study of actions and reactions of the reader, and the mental transformations necessitated by the human when presented with reactions, sometimes consistently biased, from the machine.

Furthermore the *w-graph* is written with insufficient attention to the structure of the *r-graph.* The machine interacts assuming a *model-reader*(in the image of the programmer.) Consequently, the actual reader, in his real reactions to the machine's actions, will "confuse" the machine, further contaminating proper exchange, and rendering discourse futile and insubstantial.

In this interaction, ther is no common grounds between the parties. There is no *grounding* in communication. In the next section, we will examine what the neccessary criteria are that would enable communication through an interface.

4.7 Grounding in communication

A meaningful communication is established if there is coordination in *content* and *process*, meaning that both parties agree what the subject is that they are dealing with, and the method and style they employ, respectively. To be able to establish this, Clark[Clark] writes, in *Grounding in Communication*:

> They cannot even begin to *coordinate on content* without assuming a vast amount of shared information or common ground— that is, mutual knowledge, mutual beliefs and mutual assumptions. And to *coordinate on process* , they need to update, or revise their common ground moment by moment.

In a sense, we have already touched upon the issue of *coordination of content*—we have seen that without interest on the reader's part on a particular subject, there will be no communication or interactivity.

About the *coordination of process*, it will only be successful when the two parties are aware of each other's state, manners, as well as the each other's goals. We have modelled the two parties in an interaction as two graphs. A graph includes representation of a party's character, beliefs and behavior, or, using other words, the party's states, manners, and goals. To enable communication, the two parties must be able to *coordinate content*(decide together what the semantics are), *manipulate context*(how syntax changes meaning). Furthermore, to coordinate in process also requires each party to have some reasonable assumption of the nature and form of the other party's graph. Each party has *memory* of the other party's past behaviour, and thus can gradually construct and modify an assumption-graph that models the other party's actual graph.

We will examine various points in the following subsections.

4.7.1 Coordination of content

Lewis Carroll in Alice in Wonderland:

> Alice, when talking to Humpty Dumpty, asked whether a word he used could mean what he said it meant. Humpty Dumpty: "Who is to be the master? The Word or I?"

The meaning and effect of an image is a function of cultural parameters, conventional associations and usage. An image of a tiger can signify, outside of the representation of the actual striped animal, strength, integrity or danger.

An interactive movie can use a system where an individual image, or an image sequence, can be "tagged" to signify an arbitrary set of things (cognitive objects, associative adjectives, etc.). These tags can, in the next developmental stage, be used as automatic script guiders. The script then can be as free-form as "Thunder when the Evil Witch of the West next meets Oz". This is the beginning of an intelligent script-reader like the ones implemented for computer graphics in the now-defunct Vivarium Group in the Media Lab. The movie runs off on its own.

An interactive movie should alow the reader to effect content (the connotations contained) of individual images. The *content* of an image is effected simply by supplementing the list of signifiers attached to it. The reader thus decides the content of image and its associations. It is the machine and the way it represents images that should offer the ability to assimilate new information, and, at a later time, present it in suggestive and pliable fashions when invoked. The machinery listens; and the dialogue that ensues enriches the language of the images. This implements what is actually true of the plasticity of images in general—that the *meaning of an image changes with use.*

4.7.2 Manipulation of context

The content and meaning of any imagery cannot be dealt with without considering its context. Context affects content. Different montages (structure of editing) of the same material produce entirely differently meanings. Using and arranging specific segments of *Star Wars* can make its reading more horrifying than *Halloween VI* or more romantic than *Aria.*

An interactive movie should take advantage of the possibility of allowing the reader to participate in the actual creation of the movie itself. He should be presented with *tools* [12] with which he can then change the meaning of the movie by adding (or removing) montage links, and thus transform its accents and moods.[13] As in any reasonable discourse, both parties form the manners of the exchange.

[12] Tools are software units in the form of user-friendly objects. An example would be a *linker* that prompts linking two segments producing a virtual montage point.

[13] This is computationally feasible, and in fact has been experimentally implemented by the Interactive Cinema Group at the Media Lab. Robin[Robin 90] of the Visible Language Workshop at the Media Lab has also done related work in scripting.

The reader becomes a secondary writer. He modifies the *script*.

4.7.3 Acknowledgement of intentions

A major cause of miscommunication, between two systems, is wrong assumptions and beliefs of one system of the other. Dennet [Dennet 79], in his examination of the interaction between a human and another complex system, coins the term *intentional system* as

> ...a system whose behaviour can be—at least sometimes—explained and predicted by relying on ascriptions to the system of beliefs and desires (and hope, fears, intentions, hunches,...).

Surely, the virtual writer must be designed in such a manner such as to treat its reader as such a system. The virtual writer must begin an interaction by making certain assumptions of its reader's desires, beliefs and intentions; it must give to its reader, a loan of a set of goals, desires and beliefs. In designing an interactive system, the system must assume that the reader has intentions. Most systems however, to avoid complexities, instead use a *design stance* in its view of its reader: "if he clicks on A, I must give him B." This *design stance* does not work.

For an interaction to be more substantial, the *virtual writer* must view the reader as an *intentional* system. But how do we begin to represent attributions of beliefs and desires?

4.7.4 Frames of mind

A *r-graph*, with its nodes and edges, can be used to model a set of desires, beliefs and goals.

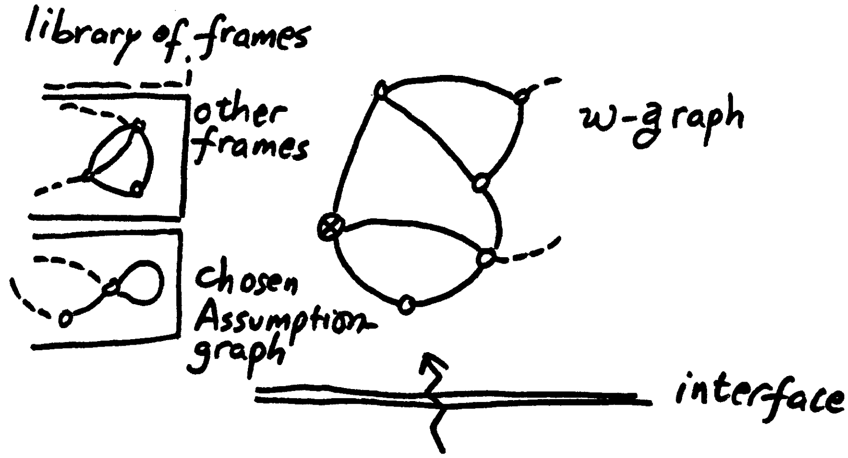


Figure 4.2: A set of assumptions of the other party is taken by choosing and modifying a *frame* from a library.

The two sides of interaction, on the side of the reader and the side of the computational structure of the machinery, are represented with the *r-graph* and the *w-graph.* For any constructive communication, both sides must begin to build an image of the other's graph, the other's "goals, desires and beliefs." But this "building" should not begin from nought. Instead, given some initial input, a large amount of assumptions should be taken which can subsequently be modified. In other words, from a library of frames (library of graphs and associated data), one *frame* should be chosen and adopted as an *assumption-graph* of the other party; and as interaction proceeds, these "goals, desires and beliefs", represented by this envisioned graph, must be modified to gain best results. It is a "tree of knowledge" that consistently needs pruning and trimmings as new data arrives. However this frame can be abandoned for another if, after some interaction, it appears to be an unsuitable one.

Minsky[Minsky 74] wrote of a good example of a person interacting with a new environment. The person walks into a room that has a table with four chairs around it. He immediately refers to a *frame* he has in his mind, and assumes it is a living room. He sees not the entirety of each chair, for they are visually blocked by the table. But he deduces what the rest of the chair to be. He has "assembled" a *graph* that represents his knowledge. As he walks around the room, he discovers a cup on the seat of one of the chairs. No panic— he attaches another node to his graph for it. But if by some insight he realizes that it is not a living room but a dining room, he must abandon this graph for a new *frame* of mind. There are *links* between associated *frames*.

4.8 Guidance in dramatic structure

Of course not all disenchantment with interactive movies stem from the lack of grounding in communication. Given that interactive movie has but a short history, there are no good models, no good experiences to draw upon. But there is a rich history of linear narration, and with it has developed many effective rules of presentation. Are there no lessons we can learn from them?

An example is dramatic structure—set-up, adventure/conflict and climax/crisis denouement. The nature of interactivity need not eliminate this. Even with input from the reader, drama can still be guided with some basic principles. This is being attempted by Joseph Bates [Bates 90] in his project *Computational Drama in Oz.*

He writes:

> ...rapidly search abstract plot space for controllable paths that have the desired dramatic structure. We can view this as a kind of abstract adversary search, where we define a set of abstract operators, means for mapping operators into concrete moves, means for recognizing the abstract effects of the users moves, and an evaluation function on event

histories (ie, stories) that lets us recognize sequences with "good" dramatic structure.

4.9 Ethnographic design

One major pitfall in interactive movie programming is that it is ethnocentric. Minorities have traditionally been rejected. It has also been shown [Spender 80] that women have been consistently ignored in texts. These and other disenfranchised groups form the majority of readers and thus it should only be expected that movies will fail if these readers are not taken as central in designing the nature of the programming.

How can this problem be approached? To start, an analysis of interfaces and programming can reveal attributes and deficiencies in ethnographic criteria. From such analyses, guiding principles can be assembled with which future programming can follow. Each new programming can in turn be evaluated and advised according to these principles.

4.10 Interface Bottleneck

Numerous problems in interaction arise from the limited channel bandwidth of the interface. A mouse as an interface gives only the options of placement in 2-D and whether or not buttons are pressed at any time. For the machinery to successfully communicate with readers, it must be able to read them, and that is not possible before a more advanced interface is developed that offers a larger bandwidth and more flexible interface channel.

Chapter 5

Currents

In this final chapter, we reiterate and re-examine the goals of representing imagery, and the motivations and idealism behind implicit coding and content-based reprsentations.

5.1 Objectivity vs. subjectivity

We pause here for a moment to examine both the views of objectivists and subjectivists about representations.

For the objectivist, things exists independantly of mind, and events occur whether or not they are observed. His thesis is, as Bunge[Bunge 74] writes:

> A representation, by definition, represents something: there can be no representation in and by itself... [it] maps (some of) the features of the [depicted]. A representation is not a part, let alone the whole, of its object... a representation, however creative, *does not create its object.* To claim that it does is to muddle things up and to make it mockery of the age long striving for objectivity, that supreme accomplishment of the creative scientist.

For the subjectivist, on the other hand, there is no physical world that exists independant of our descriptions. For him, the world is our representation or part of the latter. Representations are creations not copies.

Goodman:

> There is no such thing as *the* way the world is: the world is as many ways as it can be truly described, seen, picture, etc.[Goodman 60]

> That nature imitates art is too timid a dictum. Nature is a product of art and discourse.[Goodman 68]

In this paper we subscribed to the subjectivist stance, and abandoned the historical declared aim for objectivity. We had demonstrated, in the first chapter on **Digital Messages**, that the very format of representation itself is a function of culture; it reflects the general *episteme* of the culture.

With this, we accept the futility of the search for objective representations and surrender attempts to describe the world "*the* way the world is"), We begin to then describe the world, in images and otherwise, in "as many ways as it can be truly described". We may then describe the world and images as *we see* it. Descriptions are interpretations and we accept the impossibility of hermeneutical neutrality.

What an image *contains* depends squarely on "the way we see it". That the *content*, is central to the issue of representation, suggested that representations should be constructed upon them. This is the motivation for our search for content-based representations.

5.2 The identity of the image

5.2.1 Aesthetics of the image

One of the goals of an artist, in the act of rendering, is to transcend the physical canvas, to endow the rendering with "consciousness," to inject it with spirituality. Allegorically, the image then speaks.

If we were to have this be an aesthetic goal in digital representations of imagery, we must begin to examine procedures that might lead to it. How can a digital image begin to *have* "sense"? Surely, one of the prerequisites is that an image *knows about itself*—it knows what it contains, and its constituents know what they constitute, as much as DNA molecules know about the organism it builds. Holography too achieves this quality: every piece of a hologram contains the entire image. Illuminating just one section still "works", though less vibrantly. What fate belies the plane-image?

5.2.2 Modeling the image

How can we achieve such aesthecism for the *registered* plane-image? How can pixels be injected with soul, and transcend the being of a artifact?

If it is not possible that each pixel be a microcosm of the image, then at least each pixel can be made to connect to all others at various levels. Each pixel must be aware of its neighbors, each pixel must have the knowledge, or at least the access to this knowledge, of what larger units it constitutes. If it is not possible that each pixel contains all of the image, then we will have to be contented with the image as a constructed heterarchical structure of relationships. At least then each pixel can know the entire image 'through' other pixels.

A model representation of the *registered* plane-image then, as argued in **Methodology**, is computer graphics descriptions. These *synthetic* descriptions that begin with description of objects as identities, are based upon and reflect cognitive perceptions. In these descriptions, each pixel "knows" very well about the image and how it is constructed, for the pixel is actually the manifestation of the descriptions.

Computer graphics offer valuable guides to a cognitively structured description of *registered* image sequences.

5.2.3 Analysis vs. synthesis

One of the driving forces behind computer graphics is the attainment of realism—to imitate and simulate reality by constructing an image from symbolic descriptions of entities. A virtual universe is simulated[Baudrillard 83] through exact descriptions.

On the other end of the spectrum, our desired representation of *registered* imagery is to process and analyse according to cognitive principles, so that we can describe them synthetically.

Analysis and synthesis is merging, or at least their goals appear to be so. There is no longer any conceptual difference between the descriptions of computer graphics and *registered* imagery. Watlington's *Video Finger* is a clear example of this merge. It is in this direction that further developments of implicitly coded content-based representations should head.

5.3 The Knowledge of the Image

5.3.1 Tags, links, and information

The model computer graphics description for *registered* image sequence gives us some guidance of what we wish to achieve.

We are not at the point to achieve an implementation with substantial structural similarities, but certain desirable qualities can be achieved. The implementation in the chapter of **Methodology** incorporates some of them. *Analogies* formed *characters* that represent cognitive objects, and ordered sets of contiguous frames into *shots*. *Sympathies* set sequences into context with *links*. Short of achieving the ideal, the representation inherits a mishmash of encoded information, tags and links. It needs major study before the form of the model representation can be echoed.

Is this mishmash of encoded information worthwhile?

It is likely that the information embedded in images, and the *links* between them might be inconsequential and irrelevant. It is then likely that the entire endeavour might be discounted as fanciful and foolish. Our aim was to encode meaningful knowledge into images, to "inject the image with soul", and we end up with a possible clutter. Is the entire concept of content-based representation unrealizable?

5.3.2 Knowledge vs. information

If the image does not have knowledge, at least it has information.

In a keynote address to the 37th Annual Conference of the Canadian Library Association, Heinz Von Foerster speaks of books in libraries as sources of information that is crucial for research, he highlights the often vague distinctions between *information* and *knowledge*. Foerster speaks:

> ... The perversion of the notion of knowledge, and the related misconception of the social function of the library and its librarians is connected to a confusion that presents the library as a repository of knowledge and information. However, a library cannot store knowledge and information—only documents, books, maps, microfiche, slides, etc. When people use these materials

> they will become knowledgeable and informed. By obfuscating this distinction, knowledge and information can be made to appear as if they were commodities, to wit, the emerging "knowledge industry", "information processors," etc. With this the problems of how to know and how to let know are successfully pushed into a cognitive blind spot.[Foerster 82]

5.3.3 Knowledge through information

Knowledge is naturally the desired, the end, but without *information* there can be no such end. Only with a sufficiently large bank of information can there be deductions, excavations of relationships, coherent understanding manifested by hypothesis and thesis. Of course, although *information* is necessary, this same *information* must be organized in an accessible fashion, ordered in a manner that encourages examination. Foerster:

> ...I do not perceive a library as a bonum librorium copia, or a collection of good books, and the librarian as the custodian for these books. No! I see in a library the most awesome manifestation of the traces of our civilization and the librarian, in the role of a midwife or an obstetrician, helping people giving birth to new ideas, understanding and insights.

5.4 Concluding Remarks

In our idealism of content-based reprentation, the intent was to embed an image with knowledge. We have instead an image embedded with information. The distinction between knowledge and information is connected with the definition of content. We shall leave it at that, and persist with our contentions and take solace in claiming the fact that the image's new properties do make the the image *richer*.

Glossary

registered image An image recorded of the natural world with a taking instrument such a camera.

analogon An image array as an object.

first message Pure denotation of the depicted scene.

look-up-table List of referenced items, such as colors, of an image.

content A cognitive entity in an image.

context Relative placement of, and relationships between chunks of information.

plane-image A flat surface format of an image such as a photograph.

colormap A look-up-table for colors in an image.

vector quantization Method to find best limited set of representative vectors.

link A symbolic connection between two image sequences.

image segmentation A spatial section of an image.

character An entity derived from analogous segmentations across successive frames.

shot A visually contiguous sequence of image frames.

episteme Paradigm through which knowledge is gathered.

implicit coding Coding of information within representation.

explicit coding Coding of information alongside representation, such as the use of an alpha-channel.

Acknowledgements

My first and foremost thanks go to Walter Bender without whose advise and criticism this paper would not have been possible.

Thanks also to my readers, Glorianna Davenport, Otto Piene and Andy Lippman.

And to colleagues Janet Cahn and Julean Simon for their advice to be more formal in writing.

Thanks is also due to Panos Kouros and Thomas Smith for inspirations.

And to Patrick McLean and Peter Droege for their greatly needed moral and immoral support.

To Fred Feinberg for our correspondences and exchanges.

Special thanks for the people with whom I had stimulating interactions in my office.

And lastly to all of the Media Lab—there are many fine people here.

Bibliography

[Adelson 84] Adelson, E. et al. Pyramidal Methods in Image Processing. *RCA Engineer.* Nov/Dec 1984.

[Barthes 85] Barthes, Roland. *The responsibility of forms: the photographic message.* Hill and Wang, New York 1985.

[Barthes 77] Barthes, Roland. *Image, Music, Text: Rhetoric of the Image.* London: Fontana, 1977.

[Bates 90] Bates, Joseph. Computational Drama in Oz. *Art Com.* Dec 1990.

[Bateson 72] Bateson, Gregory. *Towards an Ecology of Mind.* Ballantine Books, New York 1972.

[Baudrillard 83] Baudrillard, Jean. *Simulations.* Semiotext(e) 1983.

[Bernheimer 61] Bernheimer, Richard. *The nature of representation: A Phenomenological Inquiry.* N.Y.U. Press, New York 1961.

[Blatt 84] Blatt, Sidney. *Continuity and Change in Art: The Development of Modes of Representation.* Lawrence Erlbaum Associates, New Jersey 1984.

[Borges 57] Borges, Jorge Luis. *El Aleph.* Buenos Aires: Emece 1957.

[Bove 87] Bove, V. Michael. *Using a Media-Cut Algorithm to Simplify Lighting Calculations.* Media Lab MIT, Cambridge 1987.

[Bunge 74] Bunge, Mario. *Treatise on Basic Philosophy v.2. Semantics II: Interpretation and Truth.* D. Reidel Publishing, Dordrecht, Holland 1974.

[Clark] Clark and Brennan. *Grounding in Communication.* Stanford University 1990.

[Davenport 89] Davenport, G. and Brondmo, H. Creating and Viewing the Elastic Charles—A Hypermedia Journal. *Hypertext II Conference proceedings.* York, England, July 1989.

[Davenport 91] Davenport, G. et. al. Cinematic Primitives for Multimedia. Summitted to be published in special issue of *IEEE on Multimedia*, Summer 1991.

[Dede] Dede, C. The role of Hypertext in transforming information into knowledge. *Proceedings of the National Educational Computing Conference,* 1988 pp.99-110.

[Dennet 79] Dennet, Daniel. *Brainstorms.* Bradford Books, MIT Press, Cambridge Mass., 1979.

[Doland 89] Doland, Virginia M. Hypermedia as an interpretive act. *Cognitive Issues.* Vol. 1, No. 1. Spring 1989.

[Drexler 87] Drexler, Eric K. *Hypertext Publishing and the Evolution of Knowledge.* Draft 2. Stanford Computer Science Department.

[Eco 76] Eco, Umberto. *A Theory of Semiotics.* Bloomington: Indiana University Press 1976.

[Eco 79] Eco, Umberto. *The Role of the Reader.* Bloomington: Indiana University Press 1979.

[Foerster 82] von Foerster, Heinz. Address to 37th Annual Conference of the Canadian Library Association. *Canadian Library Journal*, Vol. 39, No. 5, 1982.

[Foerster 61] von Foerster, Heinz. Principles of self-organization. *Transactions of the U. of Illinois Symp. on Self-Organization,* 1961.

[Foucault 74] Foucault, Michel. *The Order of Things: an archaeology of the human sciences.* Tavistock Publications 1974.

[Gagnon 86] Gagnon, D. *Interactive Versus Observational Media: The Influence of User Control and Cognitive Styles on Spatial Learning.* M.I.T Audience Research and Harvard Graduate School of Education 1986.

[Gogh 67] van Gogh, Vincent. *The letters of Vincent van Gogh.* ed. by Mark Roskill. Atheneum, New York 1967.

[Goodman 60] Goodman, N. *The way the world is.* Rev. Metaphys 1960, Vol 14: pp. 48-56.

[Goodman 68] Goodman, N. *The Languages of Art.* Indiana and New York: Bobbs Merrill Co. 1968.

[Gray 84] Gray, Robert M. Vector Quantization. *IEEE ASSP Magazine*, April 1984.

[Grice 75] Grice, H.P. Logic and conversation. In P. Cole and J.L.Morgan (Eds.): *Syntax and semantics, Vol 9: Pragmatics.* New York: Academic Press 1975.

[Haase 90] Haase, K. *Notes on Knowledge Representation.* MIT. Cambridge, Mass.

[Higgins 67] Higgins, David. *Identity and Spatio-Temporal continuity.* Oxford, Blackwell 1967.

[Leonardo 56] Leonardo da Vinci. *Treatise on Painting.* Princeton University Press 1956.

[Marr 82] Marr, David. *Vision: A computational investigation into the human representation and processing of visual information.* W.H. Freeman and Company, N.Y. 1982.

[Metz 74] Metz, Christian. *Semiotics of the cinema.* Oxford U. Press 1974.

[Minsky 74] Minsky, Marvin. *A Framework for Representing Knowledge.* MIT-AI Lab Memo 306, June 1974.

[Moholy-Nagy 65] Moholy-Nagy, L. *Vision in Motion.* Paul Theobald and Co., Chicago 1965.

[Mondriaan 20] Mondriaan, Pieter. *Le neo-plasticisme.* Paris, Eidition de l'Effort moderne 1920.

[Netraveli 88] Netravali, A. and Haskell, B. *Digital Pictures: Representation and Compression.* Plenum Press 1988.

[Olson 74] Olson, D.R. and Bruner, J.S. Learning through experience and learning through media. In D.R.Olson (Ed), *Media and Symbols: The Forms of Expression, Communication and Education.* 73rd Yearbook of the National Society for the Study of Education. Chicago. University of Chicago Press 1974.

[Panofsky] Panofsky, Erwin. *Prospettiva come forma simbolica.* Milano: Fettrirelli.

[Piaget 71] Piaget, Jean. *Genetic Epistemology* (E.Duckworth, Trans.) New York: Norton, 1971.

[Putnam 88] Putnam, Hilary. *Representation and Reality.* Cambridge, Mass. MIT Press 1988.

[Robin 90] Robin, Laura. *Multimedia Scripting.* Master's Thesis, Visible Language Workshop, Media Lab, MIT 1990.

[Simon 81] Simon, Herbert. *Sciences of the Artificial.* MIT Press 1981.

[Sontag 66] Sontag, Susan. *Against Interpretation.* Farrar, Straus and Giroux. New York 1966.

[Spender 80] Spender, Dale. *Man Made Language.* Routledge and Kegan Paul, publishers, 2nd ed. 1980.

[Watlington 89] Wadlington, John. *Synthetic Movies.* S.M Thesis MIT 1989